

More drugs, less protection

A study of the impact of the system of intellectual property rights on drug development has highlighted problems that must be addressed. An influential committee meeting this week has the power to bring about change, and should do so.

There is an urgent need to establish the basis of a new balance between the availability of drugs in the poorest countries, in which millions are dying, and protecting the investments of those who developed those drugs. Members of the Trade-Related Aspects of Intellectual Property Rights (TRIPS) council meeting at the World Trade Organization (WTO) in Geneva this week face this challenge. They do so with the less-than-comfortable conclusions of a UK government-commissioned report ringing in their ears, highlighting as it does the problems in the present system.

Many of the messages in the report are not new. The added significance lies in where it comes from. It was commissioned by Clare Short, UK secretary of state for international development, from a group of six international experts led by John Barton, professor of law at Stanford University, California, and including representatives from India and Argentina. It sends a strong signal to developed countries that the 'one size fits all' approach to intellectual property rights can have a negative impact on developing countries (see http://www.iprcommission.org/graphic/documents/final_report.htm).

The report addresses the diverse roles of intellectual property rights in poorer countries' development: some have well-established research infrastructures, whereas others have valuable genetic resources and traditional knowledge. In the poorest nations, with millions of people dying from endemic diseases, the immediate priority is not intellectual property but saving lives by getting access to essential drugs. The WTO declaration on TRIPS and public health adopted last November recognizes this and emphasizes that all countries should be able to use the flexibility in the TRIPS agreement to obtain cheap versions of essential drugs. This week's council meeting will be asking countries to come to an agreement over an exemption clause that would allow countries without manufacturing capacity

to issue a compulsory licence to another country in order to import generic and relatively cheap drugs.

Compulsory licensing has successfully been used as a threat to entice the pharmaceutical industry to lower its prices, as witnessed with anti-retroviral drugs in Brazil. Even the United States, one of the most vehement opponents of compulsory licensing around the WTO negotiating table in the past, was recently vaunting its persuasive powers in forcing Bayer to reduce the price of its anthrax drug Cipro during last year's anthrax attacks. But unless agreement is reached on the exemption clause, suppliers of generic versions based in countries yet to implement the TRIPS framework, such as India, will no longer be able to supply them once it comes into force, removing the desired competition with patented pharmaceutical products.

The report calls for developing countries to make maximum use of compulsory licensing in their own legislation, but notes that the economics of supply to one country with a limited market may not attract generic suppliers. It advocates that the World Health Organization or the United Nations Children's Fund, for example, should broker a deal for a generic version of an essential drug to be supplied to a group of countries with a similar public-health need.

The Commission on Intellectual Property Rights concluded that for most developing countries, rapid economic growth is more often associated with weaker intellectual property protection than that mandated under TRIPS. But if things are going to change for the better, developing countries must be given the opportunity to participate in reform discussions, instead of leaving this to the non-governmental organizations. Only then will the playing field begin to level out. The Rockefeller Foundation's plan to launch a programme in November to give experts from developing countries and groups of indigenous people a forum to discuss these issues should be welcomed. ■

Improving proto-scientists' summers

Parents need to keep offspring occupied in the summer. Universities can help them — and brighten the future of science, too.

Suppose you have a talented 16-year-old daughter who is interested enough in science to think about studying it at university, but is not obsessed. Her school does its best to sustain her interest, perhaps with help of career materials supplied by learned societies and employers. She has taken on board the statistics showing that graduates in all sciences have very good employment prospects. But she could yet be diverted by the financial blandishments of law or accountancy, or the human intrigues of history or politics.

Maybe the next summer break will provide an opportunity to stimulate her interest. You go to Google. You'll find summer camps in the United States to make your offspring physically and/or morally healthier, or a cyber expert. You can send her to NASA space camp in Alabama, where she'll learn to fly a (virtual) space shuttle. You find US universities and government labs reaching out to seize her soul. Other countries, however, are invisible. And yet the cry the world over is that there are too few young souls knocking on university doors. Non-US universities should do much more to capture them.

Academics will hate the idea of compromising their precious summer months' research opportunities any further, but maybe they have no choice. To be fair, there are many outreach schemes already in place: open days, postdocs assisting in schools and museums. But too few universities have grabbed the idea that youngsters could be stimulated by exposing them to science departments' teaching and research activities in more depth.

More universities should invite older school students to apply for a week in residence where they can be introduced to the leading edges of research and to the diversity of training and education that they can expect to receive. Those students who show particular aptitude should be given the chance to spend a few summer weeks working alongside students and postdocs, helping with tasks that may be menial to the latter but that give young people enough of a glimpse of the world of science to be enticed further into it. And those universities enlightened enough to do this should advertise their wares in a search-friendly form on the web, to be eagerly seized by relieved parents. ■



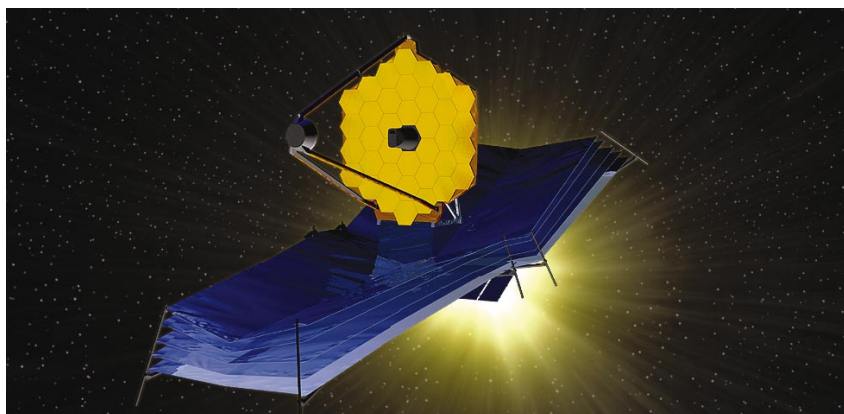
Next-generation space telescope sets course for 2010 launch

Tony Reichhardt, Washington

It has a basic design, a team of scientists to oversee it and a new name. Now one of astronomy's most challenging projects is poised to begin development.

NASA announced last week that the Next Generation Space Telescope will be named after James Webb, the administrator who led the agency during its glory days of the 1960s, which culminated in the Moon landing of July 1969. The James Webb Space Telescope, which is designed to continue operating for at least five years after its launch in 2010, is expected to cost \$2.8 billion.

The telescope, which will have a diameter of 6 metres and be optimized to view infrared light, topped the US astronomy community's ten-year wish-list of projects released in 2000 (see *Nature* **405**, 381; 2000). A large light-collecting area and sensitive detectors will enable it to see galaxies that are fainter,



Petal power: an artist's impression of the new telescope, whose mirror will open like a flower.

older and farther away than any seen by the Hubble Space Telescope. Astronomers are hoping they will be able to learn how large

structures formed in the early Universe, and to observe the birth of stars and planets with unprecedented clarity.

This won't be easy. The design incorporates several untested ideas, including the first placement of a large observatory at the L2 point, 1.5 million kilometres away from Earth in the opposite direction to the Sun. As a result of the combined influence of the gravitational attraction of the Sun and Earth, objects stationed at L2 orbit the Sun once a year and remain in roughly the same position relative to Earth.

L2 is colder than Hubble's orbit around Earth, so the Webb can cool itself using a sunshield instead of bulky and expensive mechanical systems. Stationing the telescope at L2 also makes it easier to target than Hubble, which has to avoid pointing at the Earth or Moon. This should bring the operating costs of the new telescope down to around a quarter of those required for Hubble, says project manager Bernard Seery of NASA's Goddard Space Flight Center.

L2 does have one main disadvantage: a telescope stationed there cannot be repaired or upgraded by astronauts, as Hubble can. If a crucial onboard system fails and cannot be fixed by radio command, the mission is likely to end.

The winning design for the telescope — which was put forward by an industrial consortium led by TRW Space and Electronics

US return may boost 'S' in UNESCO

Jonathan Knight

The United States has promised to rejoin the United Nations Educational, Scientific and Cultural Organization (UNESCO), nearly two decades after it severed ties with the organization.

President George W. Bush made the announcement during a speech to the UN General Assembly in New York on 12 September, saying that the organization had been sufficiently reformed. The United States withdrew its funding and staff in 1984 on the grounds that UNESCO had become too bureaucratic and politicized.

UNESCO, which was founded after the Second World War to promote international collaboration on scientific and cultural issues, has a relatively small science budget of about US\$50 million. It runs several influential international science bodies, such as the Intergovernmental Oceanographic Commission, which helps to coordinate ocean research.

The decision was welcomed by US

scientists who had urged their country to rejoin the organization.

"It seemed rather anomalous that a great country like this wouldn't support it," says James Cronin, a physicist at the University of Chicago and a member of the Pierre Auger Observatory, a cosmic-ray detector currently under construction in Argentina (see *Nature* **419**, 12–14; 2002) that UNESCO has helped to fund. Cronin was also one of 37 Nobel laureates who unsuccessfully petitioned President Bill Clinton to rejoin the organization in 1993.

The impact of the decision remains unclear, however. UNESCO's budget will not immediately increase, as the United States' contribution of \$60 million a year will be subtracted from the total payments made by other countries. But the move could allow US researchers to push for the organization to divert more of its resources to scientific projects.

"The 'S' stands for science, but it's the smallest 'S' you can imagine," says Cronin. ■

of Redondo Beach, California and Ball Aerospace of Boulder, Colorado — borrows from research on large space mirrors developed for military satellites. The segmented primary mirror divides into three panels, which, once in space, will unfold like flower petals. The technical difficulty of testing this system is the main reason for a slip in the launch date from 2008 to 2010.

Once in space, the mirror, which will be made of glass or beryllium, will be tested and adjusted every few weeks to compensate for any changes in shape caused, for example, by heating.

These and other new technologies have pushed up costs, however. Last year, astronomers reluctantly agreed to NASA's proposal to decrease the size of the mirror from 8 to 6 metres, reducing the telescope's resolving power by 25% and its light-collecting ability by 44%. An orbital test of a scaled-down mirror was also scrapped.

Some outside observers worry that the current budget is still too small, but Seery says his team is far more savvy about the technological challenges and true costs than it was.

Others say that it is right to push the state of the art forward with the planned new telescope.

"These kinds of telescopes only come along once every decade or so," says Jonathan Lunine, a planetary scientist at the University of Arizona in Tucson and a member of the Webb's recently appointed Science Working Group, which will meet for the first time next week. "Because of the size and the scope of this mission, the opportunity to do new technology is really there."

http://ngst.gsfc.nasa.gov

Prion research stepped up as fear grows of deer disease

Rex Dalton and Erika Check

As concern mounts in the western United States about a relative of mad cow disease found among deer and elk, federal agencies are boosting research into this family of fatal neurodegenerative conditions.

Awareness of chronic wasting disease (CWD) has risen in the past few years. The disease, present in the Rocky Mountain states since the 1960s, has recently been found in the midwest (see *Nature* **416**, 569–570; 2002). Officials are also concerned that mad cow disease might enter the United States.

CWD, mad-cow disease and the human form, Creutzfeldt–Jakob disease (CJD), are all thought to be caused by infectious prions — misshapen forms of a protein that can convert the normal version to their own rogue shape.

There is no evidence that CWD has caused human disease, but this remains a possibility as 60% of people in the affected region eat venison or elk. Fears about CWD have been fuelled by reports of venison-eaters who have died with neurological symptoms, although none was found to have had a prion disease.

The Department of Defense is leading the push for more research with its National Prion Research Program. The Institute of Medicine held a meeting in Washington on 12–13 September to advise on priorities for the first grants, worth \$42.5 million.

High on the agenda was the development of better diagnostic tests that could reveal infection before symptoms emerge. Cases of prion disease are currently confirmed



Waste concern: can disease pass to humans from animals such as this infected white-tailed deer?

by taking biopsies of tonsils or brain tissue.

The Department of Defense also wants to award grants to increase the size of the small US prion-research community. Researchers are put off by biosafety requirements and may be fearful of working with prions, says Pierluigi Gambetti of Case Western Reserve University in Cleveland, Ohio. "The reality is that you have to pay people more to work in your laboratories," he says.

Gambetti heads the National Prion Disease Pathology Surveillance Center, which examines cases of neurological disease and maintains CJD tissue samples. Its budget is to rise fourfold to \$1 million next year.

In addition, the National Institutes of Health wants to create a repository for reagents used in prion research. It is also considering creating a central repository for antibodies, which are necessary for the development of diagnostic tests.

AP PHOTO/COLORADO DIVISION OF WILDLIFE

Japan celebrates safe launch after string of problems

David Cyranoski, Tokyo

The successful launch of Japan's H-IIA rocket last week provided a welcome boost to the country's troubled rocket programme, before a crucial international launch later this year.

On 10 September, the National Space Development Agency of Japan (NASDA) used H-IIA to launch satellites designed to conduct data-relay and zero-gravity research. The H-IIA series and its predecessor, the H-II, have had several failures. The most recent occurred in February, when the rocket failed to release a ¥600-million (US\$4.9-million) satellite designed to measure the impact of high temperatures during re-entry.

The next H-IIA launch, scheduled for

later this year, will be a major test. It will carry the Advanced Earth Observing Satellite II, with detectors made by NASDA, NASA and the CNES, France's national space agency. They will monitor chlorophyll, water vapour, sea surface temperature, sea ice and ocean wind velocity.

The mission will also include Australia's first scientific satellite launch for 30 years. FedSat, designed by Australia's Cooperative Research Centre for Satellite Systems, will carry out engineering, communication and science experiments.

NASDA hopes that more successes will help to establish Japan's space programme as commercially viable. It plans to turn over operations of the H-IIA to Mitsubishi Heavy Industries in the next two years.

NASDA



What a blast: last week's successful launch could put Japan's space programme back on track.

Kosovo's ethnic divide blights UN science rebuilding plans

Quirin Schiermeier, Munich

Attempts by the United Nations (UN) to reconstruct science and education in Kosovo may be ending prematurely, according to UN officials working in the area.

Ethnic tension and corruption in the region mean that it is too early to hand over responsibility to the local self-administration, says Michael Daxner, who sets science and education policy at the United Nations Interim Administration Mission in Kosovo. But with some powers already transferred to the new administration, which Daxner claims is discriminating against the country's Serb minority, he says there is little more he can do, and last week he announced plans to leave his post by the end of the month.

The interim UN administration was set up in June 1999 after NATO bombing forced former Serbian president Slobodan Milosevic to agree to a peace plan hammered out by the G8 group of leading industrialized nations. Education and science were seen as key elements in the bid to rebuild the country's infrastructure and bring reconciliation between the Albanian-speaking majority and the small remaining Serb minority.

When Daxner and his colleagues began work in spring 2000, they found many schools and laboratories destroyed, and a climate of open violence. "Our main concern in the emergency period after the war was really to keep people from killing one another," he says.

The University of Pristina, then the region's only university, also bore scars from the recent conflict. Milosevic had eradicated the Albanian language from the formerly bilingual institution, and many Albanian-speaking professors had been expelled.

"When our premises were occupied by Serbian staff, some of us continued teaching in private houses, while others went to labs in the Albanian capital Tirana, or to Turkey or Germany," recalls Rexhep Ismajli, a linguist and secretary-general of the Academy of Science and Art in Kosovo.

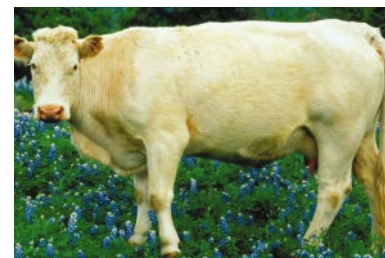
Thanks to aid of around 120 million euros (US\$116 million), roughly half of it donated by the UN, a makeshift education system has been set up. Primary-school education is in place throughout much of the region. And the University of Pristina, now one of two higher-education institutes in Kosovo, hosts some 25,000 students. But equipment and Internet connections are scarce, and there are few opportunities for scientific research.

Since the war, the university has also been dominated by an anti-Serbian ideology. Serbian students can enrol on courses, but security issues and discrimination make it impossible for them to study there. "Kosovo's Albanian-speaking majority is still extremely prejudiced against the Serbian minority, and the regional self-administration tends to cement the ethnic divide," says Daxner, a sociologist and former president of the University of Oldenburg in Germany.

Attempts by the UN interim administration to assist a new Serbian-language university in Mitrovica, north of Pristina, have also run into difficulties. The fledgling institution is being backed by the government in neighbouring Serbia, which is providing funding and staff. But many Kosovans view the idea as unwanted political interference. "We are not denying the right of Serbs to get higher education, but any mix of competencies between Pristina and Belgrade has the potential for conflict," says Ismajli. "Some people at the interim administration have the biased idea that Kosovo is a wasteland, lacking academic tradition. We are grateful for the UN's help in many ways, but there are many things which I believe we could have done better alone."

Daxner says the UN policy of handing over power to the local administration, the implementation of which began earlier this year, meant that he could make no headway on the Mitrovica university. He raised these and other problems, including nepotism, in a report presented to the Kosovo Education Conference, held in Pristina this June by the Organisation for Economic Co-operation and Development. But with the Kosovo authorities denying his charges, and further transfers of power planned, Daxner says there is little scope for him to change the situation.

Second round of gene sequencing goes down to the farm



Milking time: the cow's genome may be a new route to major mammalian genes.

Kendall Powell, Washington

The cow, dog and a single-celled genetic powerhouse called *Oxytricha trifallax* are the latest organisms to be named as high priorities for genome sequencing by the US National Human Genome Research Institute in Bethesda, Maryland.

On 11 September, the institute announced its second round of sequencing priorities, having asked researchers to submit 'white papers' making the case for their favourite organisms. The newcomers join the chicken, chimpanzee, honeybee, the sea urchin *Strongylocentrotus purpuratus*, the protozoan *Tetrahymena thermophila* and 15 species of fungi (see *Nature* 417, 473–474; 2002).

Steven Kappes, director of the US Department of Agriculture's Meat Animal Research Center in Clay Center, Nebraska, says that the cow offers "another avenue" to identify important mammalian genes. Whereas biomedical researchers use lab animals primarily as models of human disease, agricultural scientists tend to focus on normal processes such as muscle development and reproduction.

Man's best friend, however, shares many of our maladies. The dog's supporters touted the animal as a model for various diseases, including cancer, heart disease and deafness.

The protozoan *Oxytricha*, meanwhile, is an evolutionary oddity. It seems to package each of its 30,000 genes into a tiny individual chromosome. Each time it divides, it spits out accumulated 'junk' DNA. As a result, *Oxytricha* crams about the same number of genes as the mouse into a genome some 60 times smaller.

Oxytricha had previously been ranked as a moderate priority, alongside the rhesus macaque — to the bitter disappointment of researchers who use this monkey in biomedical research. The macaque genome proposal will be resubmitted in October.



Language barrier: Kosovo's new university in Mitrovica receives funding from Serbia.

India shuts door on embryonic export market

K. S. Jayaraman, New Delhi

Fears that India's booming assisted-fertility industry could become a source of human embryonic stem (ES) cells have prompted the country's medical authorities to propose a ban on the export of all human embryonic material.

The Indian Council of Medical Research (ICMR), a government-funded body that coordinates biomedical research, last week released new draft guidelines aimed at preventing the misuse of human embryos stored at the hundreds of unregulated fertility clinics that have sprung up in India over the past ten years. The new rules will prohibit the export of any part of an embryo, and stipulate that couples may only donate, and not sell, embryos to researchers inside India.

T. C. Anandkumar, research director of the Hope Fertility Clinic in Bangalore and a former ICMR director, was involved in drafting the guidelines. He says that the council has reacted to fears that restrictive embryo-research regulation in Europe and the United States could force scientists to look abroad for embryonic material. Some Indian clinics say that they have already been approached by Western researchers wishing to obtain human ES cells.

Whether European and US researchers would want to risk the bad publicity associated with obtaining stem cells from an unregulated clinic in a developing country is unclear. Germany allows research to be carried out only on imported ES cells, but these must be obtained from certain approved cell lines. And US researchers, who are free to obtain stem cells from fertility clinics as long as their work is not federally funded, say that they have no problem in obtaining embryonic material.

The new guidelines could, however, cause problems for Indian researchers. "Prohibiting the transfer of embryos could virtually prevent us from collaborating with foreign agencies," says Mitradas Panicker, a neurobiologist at the National Centre for Biological Sciences in Bangalore, one of two Indian laboratories that the US National Institutes of Health lists as providing ES-cell lines with which federally funded researchers are permitted to work (see *Nature* 412, 665; 2001). Panicker says that collaborations with US scientists would suffer "if the guidelines failed to make a distinction between selling embryos for money and transferring them for research".

The owner of the other ES cells on the NIH list, Reliance Life Science in Mumbai, faces similar problems. The company says that it has developed a cell line that differentiates into insulin-producing pancreatic cells that could be used to treat diabetes, and is ready to begin animal trials. "We have several



Stem-cell researcher Satish Totey fears that a ban on embryo exports may harm collaborative projects.

foreign collaborations planned with our cell lines, but as the guidelines stand today we cannot do any collaborative research," says Satish Totey, the company's head of ES-cell research.

The guidelines may yet be refined before a

final version is presented to the government. The National Accreditation Authority, a new body that is currently being set up to administer the guidelines, says that it will consider suggested amendments from the scientific community until the end of December. ■

Telescope to track speedy satellites

Geoff Brumfiel, Washington

Plans for a new military satellite-tracking telescope could help astronomers spot unusual changes in the night sky.

The 3-metre Space Surveillance Telescope (SST), a US\$65-million project overseen by the Defense Advanced Research Projects Agency (DARPA), part of the Pentagon, is designed to improve the US Air Force's ability to monitor satellites and orbiting space debris. Work on the telescope's detectors is set to begin in October and the telescope should be complete by 2008.

Building a telescope that can scan the sky rapidly for fast-moving satellites will

require some unusual technology, according to Timothy Grayson, who oversees the SST project at DARPA's Special Projects Office in Arlington, Virginia. Most telescopes use two mirrors to focus light onto detectors, but the SST will add an extra mirror to reduce distortion around the edge of its images, effectively widening the field of view. The three-mirror system will be difficult to set up, says Grayson. The design will also require the development of new, curved charge-coupled device detectors to capture digital images from the telescope.

The telescope's primary client is likely to be the military, but some data may be made available to researchers. The broad field of view and sensitive detector could, for example, help astronomers search for stars of variable brightness. Grayson has met officials from NASA and the National Science Foundation (NSF) to discuss how the telescope could benefit future research.

Civilian use of military telescopes is not unprecedented, says Jim Breckinridge, programme director for Advanced Technologies and Instrumentation at the NSF's Division of Astronomical Sciences in Arlington, Virginia. His office has an agreement with the Air Force Office of Scientific Research that allows scientists to use the Air Force's 3.67-metre Advanced Electro-Optical System telescope in Hawaii. ■



Pyramid find reopens lost chapter of history

Alison Abbott, Cairo

The site of a lost pyramid that has eluded Egyptologists for over a century has been unearthed by German archaeologists. It belongs to Nub-Kheper-Re Intef, a Pharaoh from Egypt's seventeenth dynasty, and is the first from this era to have been found by archaeologists.

The pyramid, made of mud bricks, has collapsed, but decorations and pottery fragments in the underlying chamber will help archaeologists shed light on Egypt in the sixteenth century BC. The period preceded the golden age of the Pharaohs epitomized by Tutankhamen of the eighteenth dynasty and by the Ramesses kings of the nineteenth and twentieth dynasties. "It is very important that the pyramid has turned up," says John Taylor of the British Museum's Department of Ancient Egypt and Sudan. "It will help us to clarify the chronology of the royals and understand how they were related."

The location of seventeenth-dynasty pyramids has been pieced together over the past century from ancient documents and nineteenth-century reports from European travellers. One document, purchased by the British Museum in the 1850s, details an inspection of royal tombs carried out during the twentieth dynasty in 1115 BC after rumours of tomb raids. The papyrus records that Nub-Kheper-Re's pyramid was subject to an unsuccessful break-in attempt.

The chamber was briefly explored in 1827



The site of Nub-Kheper-Re Intef's ruined pyramid, and (inset) an older bust from his burial chamber.

but was not studied in detail. At the time, European diplomats routinely had permission from the Egyptian government to search for treasure, and paid locals to strip tombs.

Daniel Polz, a director of the German Institute of Archaeology in Cairo and head of the excavation team, relocated the site, which is on the west bank of the Luxor at the entrance to the Valley of the Kings, with the help of locals who are skilled in recognizing man-made disturbances in a desert landscape. One of the first tombs to be opened was a small chapel of a top official of the king, called Teti. The walls of the chamber are decorated and bear the name of the king in large hieroglyphs.

Nub-Kheper-Re's neighbouring burial chamber is much larger and undecorated, and its deep sarcophagus is formed from a bay cut

into the rock — an unusual feature as such sarcophagi are typically free-standing. The team also found evidence that two obelisks may have been mounted in front of the pyramid, a further unique feature. The king's chamber also contained a life-sized sandstone head of another king, probably belonging to a statue from an eleventh-dynasty temple but borrowed during the poorer seventeenth dynasty and placed in or near the pyramid.

Polz expects the complete excavation of the pyramid and its surroundings to take around five years, and is confident that traces of other seventeenth-dynasty pyramids will be unearthed. "For the first time we have information on the architecture of a seventeenth-dynasty tomb, and the combinations of features are unique," he says.

'Unusual forces' are pushing journal market off course

Sally Goodman

Libraries may be paying too high a price for scientific, technical and medical (STM) journals because the market lacks normal competitive forces, according to a report from the UK's Office of Fair Trading (OFT).

The OFT's consultation of libraries, publishers and scientific societies, published on 9 September, raised concerns that the prices of STM publications are rising faster than inflation. The large disparity in prices between commercial and non-profit journals was also noted. In 1995, the average price of a commercial STM journal was US\$485; that of scientific-society journals was US\$229.

The market is skewed, says the OFT, partly because STM journals tend to compete on quality rather than price. Cheaper journals are thus often sacrificed so that libraries with limited budgets can subscribe to well-regarded yet expensive rivals. 'Bundling',

whereby publishers give discounts to libraries that provide electronic access to all or most of their journals, also distorts the market. The practice favours publishers with large portfolios, as librarians say it is often uneconomical to cancel individual journals in the bundle and replace them with more popular ones from another publisher.

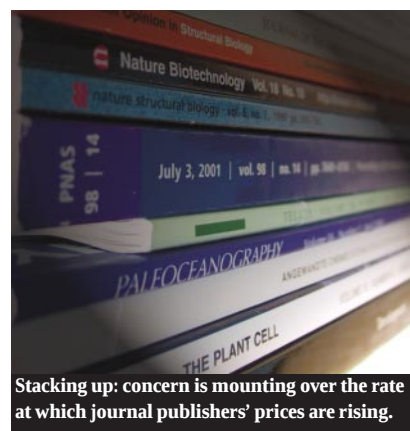
But the OFT says that intervention in the market is unnecessary at present, arguing that schemes such as the Public Library of Science, a group of academics who aim to publish free-to-access journals (see *Nature* 412, 469; 2001), could be a force for change in the future. "These emerging market forces are clear evidence that the market is already working," agrees Derk Haank, chairman of Elsevier Science, the largest STM publisher.

David Prosser, director of Oxford-based SPARC Europe, an alliance of European research libraries and institutions, says he is

glad the OFT has recognized the problem.

"But it would have been useful for them to say something a little stronger," he adds.

www.oft.gov.uk



Stacking up: concern is mounting over the rate at which journal publishers' prices are rising.

Mosquito takes flight to spread West Nile virus to California

San Francisco Health officials in California are continuing to investigate how the West Nile virus arrived on the US west coast, after tests confirmed that a Los Angeles woman had become infected.

The case of encephalitis, from which the woman has recovered, was reported as a probable West Nile virus infection on 6 September (see *Nature* **419**, 102; 2002). But health officials are puzzled as to how she became infected. No other cases have turned up west of the Rocky Mountains. Nor has close monitoring of birds and mosquitoes revealed the virus to be present in California.

Some experts believe that the woman, who lives near Los Angeles airport, may have been bitten by a mosquito that arrived on an aircraft. Mosquito-borne diseases are known to travel by plane. Six cases of malaria appeared near Brussels airport in 1995, and another four turned up around the Charles de Gaulle airport outside Paris in 1999.

NIH names heads of alcohol and mental-health institutes

Washington A leadership void at the National Institutes of Health (NIH) was partially filled last week, as NIH director Elias Zerhouni appointed new heads for two of the agency's institutes. Thomas Insel will lead the National Institute of Mental Health, and Ting-Kai Li will take charge at the National Institute on Alcohol Abuse and Alcoholism. Both are due to take over in November.

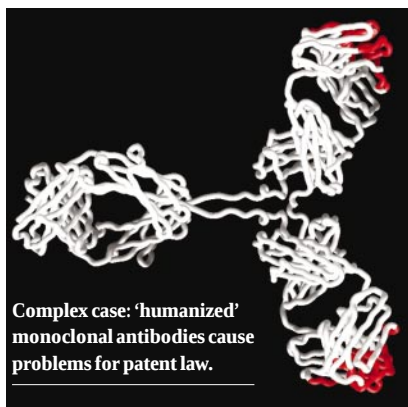
Insel is currently director of the Center for Behavioral Neuroscience at Emory University School of Medicine in Atlanta, Georgia, where he studies the role of the neuropeptide oxytocin in social behaviour. Li, director of the Indiana Alcohol Research Center at Indiana University School of Medicine in Indianapolis, works on alcohol metabolism and the genetics of alcoholism.

Li and Insel are the first new institute directors to be appointed by Zerhouni, who needed to find chiefs for several institutes when he arrived in May.

Genentech wins first round of patent battle

San Francisco In a case that has important implications for the burgeoning business of monoclonal-antibody pharmaceuticals, Genentech of South San Francisco has drawn first blood in a patent dispute with its rival Chiron.

Monoclonal antibodies are now emerging as highly profitable drugs (see *Nature* **417**, 584–586; 2002). Genentech earned around \$350 million in 2001 from its drug Herceptin,



Complex case: 'humanized' monoclonal antibodies cause problems for patent law.

which targets a receptor on the surface of breast-cancer cells. But Chiron, based in Emeryville, California, claimed that a patent issued in April 2000 gave it rights to all monoclonal antibodies against the receptor.

On 6 September, a jury in Sacramento accepted Genentech's argument that Chiron's patent covers only mouse monoclonal antibodies. Genentech's product is a 'humanized' monoclonal antibody, created by grafting the mouse antibody regions needed to bind to the receptor onto a human antibody.

Experts warn of further lawsuits as the patent system struggles to deal with these sophisticated drugs. "Unlike small molecules, which you can completely define, monoclonal antibodies are a more complex issue," observes immunochemist Martin Glennie of the University of Southampton, UK.

Fruitflies' eyes and shifting plates earn Balzan prizes

Milan Developmental biologist Walter Gehring and geologist Xavier Le Pichon are among this year's winners of the prestigious Balzan prizes for science and culture.

Gehring, who works at the University of Basel in Switzerland, was recognized in particular for his identification in the fruitfly *Drosophila* of *Pax6*, a gene that controls eye development in both invertebrates and vertebrates. Le Pichon, from the Collège de France in Paris, was honoured for his work on plate tectonics. Each scientist wins SFr 1 million (US\$660,000), half of which must be used to support young researchers' projects.

The Milan-based International Balzan Foundation awarded the other two prizes to Anthony Grafton, a historian of social science at Princeton University in New Jersey, and Dominique Schnapper, a sociologist at the Ecole des Hautes Etudes en Sciences Sociales in Paris, the first woman to be given the award.

The academic areas in which the Balzan prizes are awarded change each year. Next year's fields are genetics and evolution, infrared astronomy, social psychology and European history since 1900.

Italians rally to stop political appointments

Rome Some 600 researchers gathered in the Italian capital last week to protest against rumoured reforms which they claim would politicize Italy's science base. The reforms would split the main basic research organization, the CNR, by discipline into 15 departments; the head of each department would be appointed by the research ministry.

According to a draft document leaked from the research ministry at the beginning of August, each department head would be able to name the directors of institutes within the department. Alarmed at the prospect of political appointees being given such power, 2,500 researchers signed a protest petition. However, the deputy minister for research, Guido Possa, says that the leaked document is no more than a starting point for internal discussions about changes to the research landscape.

The CNR has only just completed a reform ordered by the previous government, which concentrated resources by reducing the number of institutes from 330 to 100 (see *Nature* **412**, 264–265; 2001). Under this plan, the heads of institutes are appointed by expert committees.

'New moon' identified as long-lost spacecraft

Washington Call it an unusual case of extraterrestrial lost and found. A chunk of debris from an Apollo lunar mission may have returned to Earth's orbit after more than three decades wandering about the Solar System.

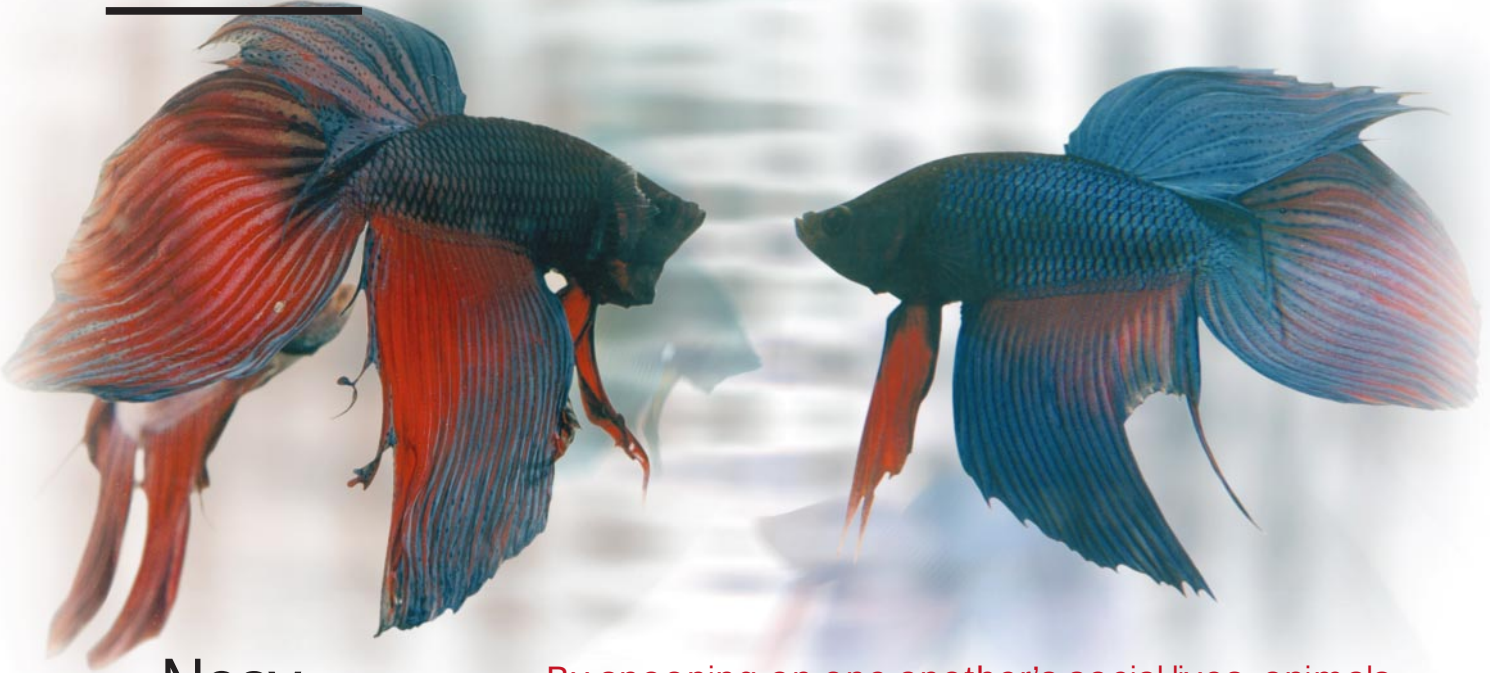
The object was discovered on 3 September by Bill Yeung, an amateur astronomer in Arizona. It was initially thought to be a third natural satellite of Earth, along with the Moon and an asteroid called Cruithne, which is some 5 kilometres across and was discovered in 1986. Cruithne moves in a complex orbit, buffeted by the Earth's and the Moon's gravity. But analysis of the new object's speed and trajectory revealed that it is more likely to be a piece of space junk.

Calculations by researchers at NASA's Jet Propulsion Laboratory in Pasadena,



Return to sender: Saturn V may have come home.

California, show that the object could be the third stage of the Apollo 12 mission's Saturn V rocket, which left Earth in 1969. If this is correct, the object would have travelled around the Sun until Earth's gravity pulled it back into orbit in April.



Nosy neighbours

By snooping on one another's social lives, animals can work out how to behave when they meet in the future. John Whitfield listens in on the natural world's eavesdroppers.

Nosiness isn't nice. But in the past few years, behavioural biologists have shown the trait in a more positive and intriguing light. Animals from fish to songbirds, they have found, can achieve success by keeping watch on their neighbours' social lives. Such eavesdropping may also be woven into the fabric of human societies — and might even help to explain why people often behave charitably.

Prying animals reap significant rewards. They know when to pick a fight and when to back down; who to mate with, and who to cuckold. Not surprisingly, perhaps, researchers have also found that animals behave differently depending on who is watching or listening. Animal communication, experts are coming to realize, has evolved to fit into a social network, rather than being a collection of signals intended simply to impress a particular mate or rival¹.

Eavesdropping shows “how incredibly subtle animal strategies are”, says evolutionary biologist Lee Dugatkin of the University of Louisville in Kentucky. This subtlety explains why it went unnoticed until recently — it's tricky to design experiments to tease out the effects on one animal of watching other animals interact. Peter McGregor, a behavioural ecologist at the University of Copenhagen in Denmark, suggests that researchers may also have neglected such experiments because they underestimated animals' cunning. “For most people, fish don't rate when it comes to cognitive abilities,” he observes.

Yet McGregor's team has shown that Siamese fighting fish (*Betta splendens*), pictured above, possess considerable social nous. Males of this famously aggressive species defend their territories with displays of fin-waving and gill-raising. But if this doesn't settle matters, things turn physical — sometimes fatally so.

Like human boxers, Siamese fighting fish study their opponents' previous bouts. Males pay more attention to their neighbours when they fight than at other times, McGregor and his colleagues found. And after viewing such contests, males approach the winners more warily than they do the losers, relying more on visual displays and less on biting².

Verbal abuse

Some researchers have questioned whether such experiments prove that bystanders scrutinize the interaction between opponents — they might be responding to the animals' inherent toughness or weediness. McGregor's team tackled this issue in great tits (*Parus major*) by using recordings of the birds' songs. To a male great tit, victory is a question of timing. A male threatens a rival by singing over his song, and shows deference by singing only in the gaps between the other's choruses. This allowed the researchers to use the same songs, regardless of any intrinsic property they might have, to denote attack or defence, belligerence or tact.

Setting up two loudspeakers outside a male's territory, the researchers played out

duets of differing structures and outcomes. They then moved either the winning or losing speaker into the bird's territory, and noted his reaction. Males sang less to losers³, perhaps because they regard them as less of a threat, or perhaps because they are more ready to escalate contests with losers to visual displays or violence. A winner got the same cautious treatment as a stranger.

Playback experiments with nightingales (*Luscinia megarhynchos*) yield similar results — except that males intensify their singing towards winners, rather than giving losers the silent treatment^{4,5}. Again, it is hard to know whether the differences reflect a more, or less, aggressive response. “The interpretation can go into hand-waving,” McGregor admits. But in each case, it is clear that eavesdropping influences the animals' subsequent behaviour.

Watching a fight also changes physiology. Cichlid fish (*Oreochromis mossambicus*) that see a contest experience a rush of testosterone, perhaps priming them to fight⁶. Dugatkin believes that the next challenge is to integrate behavioural and physiological data on eavesdropping. “There are very few studies looking at the physiology and behaviour of one system,” he says. “I think that synthesis is going to happen soon.”

Males are not the only



Nosiness alters physiology, says Lee Dugatkin.



Daniel Mennill found that female black-capped chickadees will cheat on mates that lose fights.

ones noting the results of their neighbours' squabbles — females use the same information to help them to choose their mates. Again using song playback, McGregor's team escalated contests with some male great tits, while backing down against those on neighbouring territories. Subsequently, the mates of defeated males were more likely to visit the adjacent territory⁷. Seemingly disenchanted with their partner — but impressed by what they heard coming from next door — the females were presumably seeking what behav-

ioral ecologists call 'extra-pair copulations'.

Experiments on a closely related species lend support to this idea. Daniel Mennill of Queen's University in Kingston, Ontario, and his colleagues picked playback fights with male black-capped chickadees (*Poecile atricapillus*), and then analysed the DNA of the chicks born to their mates. The researchers found that the female partners of defeated males were about five times more likely to lay eggs fertilized by other males, compared with females who never heard their partner get beaten⁸.

Covert struggle

With such high stakes, it is likely that eavesdroppers have shaped the evolution of animal communication. Some behaviours seem adapted to avoid prying ears. In many songbirds, says Mennill, the longest, most evenly matched song duels are the quietest. Where both males are struggling to dominate, he suggests, "they might not want to broadcast what's going on".

The effect of an audience on animals' social interactions is harder to study than eavesdropping, and this work is at an early stage. Again working with fighting fish, McGregor's team has found that males display to each other differently when a female is watching⁹. They reduce their aggression, and switch to conspicuous displays incorporating some of the elements used in courting, such as tail waving. And in July, Michael Kidd of the University of New Hampshire in Durham told the Animal Behavior Society's annual meeting at Indiana University in Bloomington that defeated male fighting fish prefer to court females that didn't witness their humiliation. "They have a fairly strong preference for females that didn't see them lose," says Kidd.

Eavesdropping is thought to help animals to avoid fights they cannot win. But paradoxically, eavesdroppers might make contests more aggressive, according to evolutionary biologist Rufus Johnstone of the University of Cambridge, UK. He used game theory to analyse the costs and benefits of winning and

losing fights, and of backing down quickly versus a prolonged tussle. Eavesdroppers, he found, increase the value of victory: an animal that wins its current contest will get the deterrent benefit of a tough-guy reputation, and so is more likely to escalate a fight¹⁰. "Eavesdropping can evolve to reduce the risk of fighting, but once it becomes established it promotes aggression," says Johnstone.

Replace acts of violence with ones of charity, and Johnstone's model becomes similar to those used to explain apparently selfless kindness. We often help people we are unlikely to meet again. One reason might be that good deeds get their perpetrator a glowing reputation that helps them in the future. Theoretical models suggest that altruism can survive in populations where individuals trust those they have seen cooperate with others, but give nothing to those they have seen behave selfishly¹¹.

Research by Manfred Milinski, a behavioural ecologist at the Max Planck Institute for Limnology in Plön, Germany, and his colleagues supports this idea. In one experiment, volunteers were given money and told they could donate some of it to the other participants over a series of rounds. This benefited the recipients more than the donors, because the experimenters supplemented each donation. Even though participants could not donate to someone who had given to them, they were more generous towards those who they had seen give to others¹². In another game, Milinski found that people were more likely to contribute to a public fund if their enhanced reputation could be used to attract private donations from other players¹³.

The behavioural science of eavesdropping might soon be tested in the human social marketplace. Milinski's research has attracted the attention of managers trying to control demands on Germany's health service. He suggests that doctors could publish lists of how many treatments they have prescribed and how much each has cost. Even without naming names, Milinski argues, people might be so concerned about gaining a bad reputation that they will be shamed out of seeking needless medical attention. "If peoples' reputation is at stake they are much more cooperative," he says.

John Whitfield works in Nature's news syndication team.

1. McGregor, P. K. & Peake, T. M. *Acta Ethol.* **2**, 71–81 (2000).
2. Oliveira, R. F., McGregor, P. K. & Latruffe, C. *Proc. R. Soc. Lond. B* **265**, 1045–1049 (1998).
3. Peake, T. M., Terry, A. M. R., McGregor, P. K. & Dabelsteen, T. *Proc. R. Soc. Lond. B* **268**, 1183–1187 (2001).
4. Naguib, M. & Todt, D. *Anim. Behav.* **54**, 1535–1543 (1997).
5. Naguib, M., Fichtel, C. & Todt, D. *Proc. R. Soc. Lond. B* **266**, 537–542 (1999).
6. Oliveira, R. F., Lopes, M., Carneiro, L. A. & Canário, A. V. M. *Nature* **409**, 475 (2001).
7. Otter, K. et al. *Proc. R. Soc. Lond. B* **266**, 1305–1309 (1999).
8. Mennill, D. J., Ratcliffe, L. M. & Boag, P. T. *Science* **296**, 873 (2002).
9. Doutrelant, C., McGregor, P. K. & Oliveira, R. F. *Behav. Ecol.* **12**, 283–286 (2001).
10. Johnstone, R. A. *Proc. Natl Acad. Sci. USA* **98**, 9177–9180 (2001).
11. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
12. Wedekind, C. & Milinski, M. *Science* **288**, 850–852 (2000).
13. Milinski, M., Semmann, D. & Krambeck, H.-J. *Nature* **415**, 424–426 (2002).



Good guys win: Manfred Milinski has shown that we help those with a charitable reputation.

Bridging the culture gap

Molecular biologists are deluged with data, and physicists, used to reducing complex systems to basic principles, might help to make sense of it all. But bringing the two disciplines together isn't easy, says Jonathan Knight.

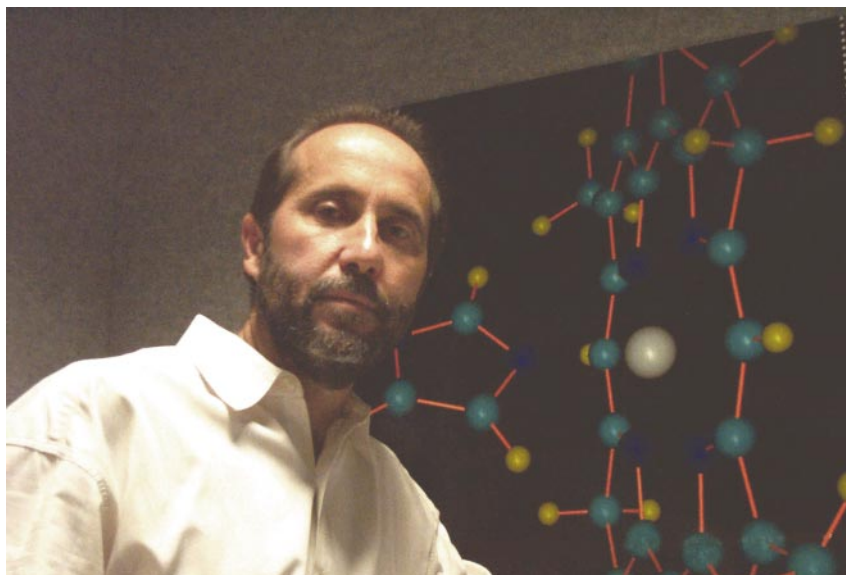
In late July, several dozen physicists with an interest in biology gathered at the Colorado mountain resort of Snowmass for a birthday celebration. Hans Frauenfelder, a physicist who began studying proteins decades ago, turned 80 this year. But unofficially, the physicists were celebrating something else — a growing feeling that their discipline's mindset will be crucial to reaping the harvest of biology's postgenomic era.

Of course, physics and its techniques have played a significant role in biology for decades. X-ray crystallography and nuclear magnetic resonance are essential tools for structural biologists. Biophysicists study everything from the forces exerted by molecular motors to the energetics of enzyme catalysis. And electrophysiologists need a working knowledge of the Nernst equation, which describes the movement of ions across cell membranes.

Many of the founders of molecular biology were also originally physicists. But in the 50 years since people such as Max Delbrück and Francis Crick created the field, it has abandoned its roots. Physics is theory-driven; molecular biology has become an empirical and descriptive science. Physics uses mathematics to represent the laws of nature; molecular biology relies on words and diagrams to describe the functions of living things. The essence of physics is to simplify, whereas molecular biology strives to tease out the smallest details. To cynics, the latter has become an exercise in molecular stamp-collecting, slotting new components and interactions into ever more complex biochemical pathways.

The two cultures might have continued to drift apart, were it not for the revolution in genomics. But thanks to a proliferation of high-throughput techniques, molecular biologists now find themselves wading through more DNA sequences and profiles of gene expression and protein production than they know what to do with. It may be time to take a step back from the details and try to see the big picture.

"Biology today is where physics was at the beginning of the twentieth century," observes José Onuchic, who is the co-director of the new Center for Theoretical Biological Physics (CTBP) at the University of California, San Diego. "It is faced with a



Deeper understanding: José Onuchic believes that physics can offer biology fundamental explanations.

lot of facts that need an explanation."

Physicists believe that they can help, bringing a strong background in theory and the modelling of complexity to nudge the study of molecules and cells in a fresh direction. "What has been all too rare in biology is the symbiosis between theory and experiment that is routine in physics," says Laura Garwin, director of research affairs at Harvard University's Bauer Center for Genomics Research, who has made her own transition to biology — she was once *Nature's* physical-sciences editor.

Opportunity knocks

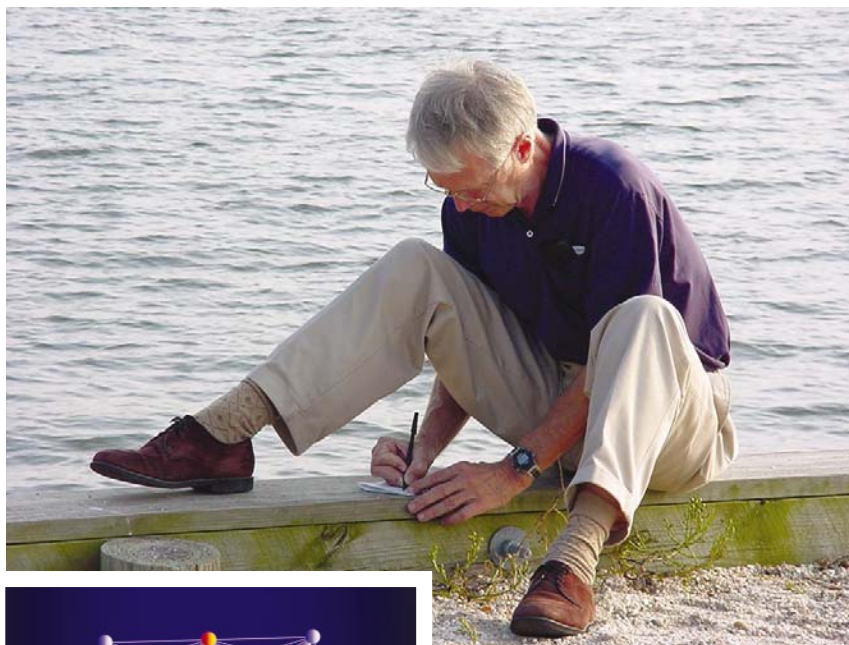
Funding agencies agree that there are real opportunities for progress in the area. In late July, for instance, the US Department of Energy awarded a \$36.6-million, five-year grant to a cross-disciplinary team at the Lawrence Berkeley National Laboratory in California. Headed by physical chemist Adam Arkin, the group plans to develop computational models of bacterial responses to stressful environments. New centres are also being established at the interface of physics and biology, and job opportunities abound. Acknowledging the trend, the American Physical Society will later this month hold a meeting in Boston — entitled

'Opportunities in Biology for Physicists' — aimed primarily at biology-curious graduate students and postdocs.

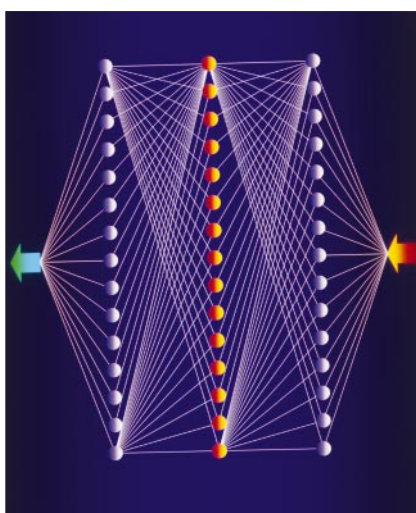
But there are pitfalls. Although some molecular biologists are happy to welcome physicists into their labs, others perceive them as interlopers who don't really understand what they are getting into. Meetings intended to bring the two disciplines together still sometimes end up with the two camps failing to communicate. And even physicists who have made the jump into molecular biology say that framing biologically relevant questions poses a huge challenge for those coming from outside.

Onuchic is in the vanguard of this new breed. Originally trained in theoretical physics in his native Brazil, the centre he now co-directs was last month given \$5.5 million over five years by the US National Science Foundation to seed collaborations between biologists and physicists. Onuchic owes his introduction to biology to physicist John Hopfield, under whom he studied for his PhD at the California Institute of Technology in Pasadena in the late 1980s, working on the theory of biological electron-transfer reactions.

Onuchic's mentor is an inspirational figure for biological physicists in general.



John Hopfield's work on neural networks (inset) showed physics can model biological processes.



Hopfield, who now works in the molecular-biology department at Princeton University in New Jersey, made his own transition to biology in the early 1980s when he developed computer models of neural networks that displayed properties of animal nervous systems. His networks consisted of virtual neurons, equivalent units that could activate their neighbours according to certain mathematical rules. Although others had been designing artificial neural networks since the 1950s, Hopfield's were the first that could recognize familiar patterns, correct errors and remember a sequence of events¹.

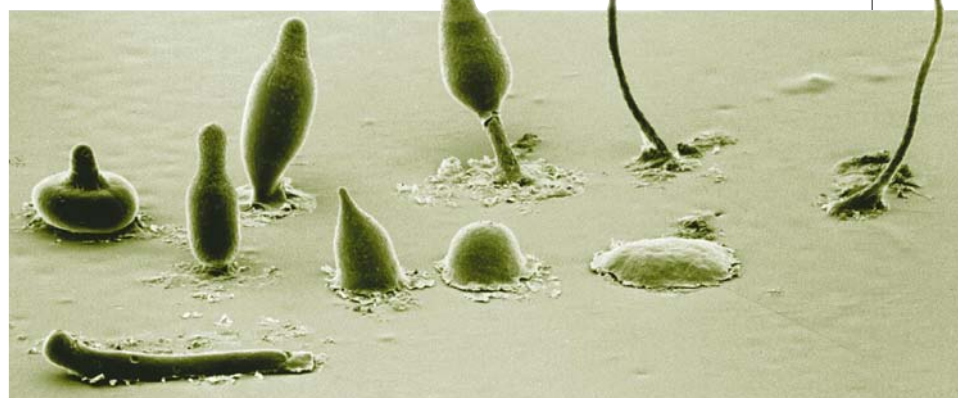
Neural networks "were a demystifying concept", says Charles Stevens, a neurobiologist at the Salk Institute for Biological Studies in La Jolla, California. They showed, for example, that complex behaviour could arise from simple repeating units, and that this behaviour could be modulated by altering the strength of the connections between the simulated neurons.

Today, physicists are exploring applications for network theory in molecular

biology. One of the projects at Onuchic's centre is investigating whether networks of gene regulation have parallels in neural networks. Just as neurons activate and repress other neurons, gene products can activate and repress other genes, directly or indirectly. The parallels may even extend to similarities between learning — which alters the firing pattern of individual neurons and hence the network's overall behaviour — and the way in which evolutionary pressures alter patterns of gene expression.

Hopfield, meanwhile, is now championing the idea of 'modular' biology. Together with cell biologist Andrew Murray, director of the Bauer centre, geneticist Leland Hartwell of the Fred Hutchinson Cancer Research Center in Seattle, and physicist Stanislas Leibler of Rockefeller University

All for one: computer models are revealing how slime-mould amoebae join forces to form these multicellular spore-bearing structures.



in New York, Hopfield has proposed that discrete biological functions rarely lie with individual genes or proteins but instead with modules comprising many interacting molecules². Examples include the ribosome, which manufactures proteins, and the signalling network of proteins that controls cell division. The researchers have suggested a number of ways to explore this idea, including efforts to reconstitute or build functional modules in the test tube, the behaviour of which would shed light on how well the underlying principles are understood.

Cracking the mould

Physicists are also helping to explain molecular influences on the behaviour of entire cells. Herbert Levine, a condensed-matter physicist at the University of California, San Diego, and co-director of the CTBP, is collaborating with biologists at Cornell University in Ithaca, New York, to model the way in which cells detect and migrate towards chemical signals. The team is focusing on the slime mould *Dictyostelium discoideum*, which lives as individual soil-living amoebae unless food becomes scarce, when the amoebae aggregate to form a multicellular organism. This produces resilient spores that hatch into new amoebae when conditions improve.

The amoebae aggregate by sending and receiving biochemical signals such as cyclic adenosine monophosphate (cAMP). But because cAMP diffuses rapidly, it isn't clear how individual cells know which direction the signal came from. Levine has devised a model that depends on the time delay from the moment a new wave of cAMP reaches one side of the cell until it completely surrounds it. The cell-surface receptors that bind to cAMP first send a repressive signal to all other receptors in the cell that prevents them from binding for about 30 seconds, Levine's model proposes³.

This basic idea did not require any special insight, Levine admits. But his background in modelling enabled him to generate

a set of quantitative predictions that can now be tested experimentally. "A biologist in principle could do the same thing, but in practice physicists have been trained to model complex systems," says Levine.

Projects such as Levine's are a promising start. But the ultimate goal for the physicists now entering the world of molecular biology is to derive fundamental principles that help to explain the characteristics of many biological systems. It is early days, but biological physicists point to Leibler's ideas about 'robustness'.

Traditionally, biologists have thought of cellular processes as exquisitely sensitive and well-balanced. But Leibler argues that cells can withstand a great deal of noise and perturbation. For example, circadian clocks, which govern organisms' daily cycles of activity, tick at the same speed regardless of temperature or the supply of nutrients. Yet the rate constants of any chemical or enzymatic process should vary with temperature, and gene expression should decline under starvation conditions. These sources of 'noise' in the system put strict theoretical limits on the type of oscillation mechanism that cells may use, Leibler argues. According to his models, a system that relied on a simple negative feedback loop to cause the production of a biochemical signal to oscillate up and down over time would not tick reliably against a noisy background, whereas one based on a fluctuating balance of positive and negative influences would be relatively insensitive to this interference⁴.

Such theoretical work should enable molecular biologists to sharpen the focus of their experiments. But in many cases, collaboration is being held back by a



Study guide: Hans Frauenfelder uses myoglobin in his work on complex matter.

The word 'function' doesn't exist in physics, but physicists need to learn about it, otherwise they will be playing in a sandbox by themselves.

John Hopfield

communication breakdown. This proved a problem in April at a workshop on DNA held at the Wellcome Trust Genome Campus in Hinxton, near Cambridge, UK. The meeting, co-organized by the Institute of Physics, Britain's main professional body for physicists, aimed to foster collaborations across the disciplinary divide. But the two camps struggled to understand one another. "The talks were too highly specialized," says Chris Phillips, a solid-state physicist at London's Imperial College who chaired a discussion session at the meeting. "A lot of physicists would have come away none the wiser. We needed an overview."

Fresh perspectives

More fundamentally, molecular biologists and physicists tend to ask different questions when presented with the same biological system. For example, when considering a network of interacting proteins, a biologist may first ask about the chain of events that occurs after one protein binds to another, whereas a physicist might want to know the rate constants of all the reactions. In a discussion session at the Snowmass meeting, a similar gulf opened up over the phenomenon of protein phosphorylation. Phosphate groups are often added or removed from proteins to modulate their activity. The biologists in the room were interested in which proteins were switched on or off in this way. But the physicists pondered a deeper question: why phosphate, as opposed to some other chemical group?

By asking different questions, physicists may in some cases bring fresh and useful insights. But one of the greatest concerns for the Snowmass participants was the risk of wasting time on issues that have no biological meaning. As several attendees pointed out, not every feature of a biological system has a functional importance, nor is every system likely to operate with optimal efficiency. Evolution doesn't choose the best option, just the best available option. Physicists stumbling into this minefield risk wasting their time on apparently interesting but ultimately trivial questions.

It's a serious concern, because there are

still lingering doubts about the relevance of earlier forays by physicists into biology. Frauenfelder was feted at Snowmass for his work on the energetics of protein folding—he developed the concept of the 'energy landscape', in which valleys represent the stable forms of the protein and hills are barriers to changing shapes⁵. But many biophysicists question whether Frauenfelder's work on the oxygen-transport protein myoglobin will lead to useful generalities about protein folding, as the principles are based on this single example. And for all their power, it remains unclear whether neural networks are genuinely mimicking the nervous system. "We still don't know to what extent they have anything to do with life," says Stevens.

Frauenfelder is unfazed by such concerns, as he is interested in myoglobin as a system in which to study the behaviour of complex matter. He likes to quote what Stanislaw Ulam, the mathematician who helped to develop the hydrogen bomb, once said to him: "Ask not what physics can do for biology, ask what biology can do for physics."

But Hopfield and others argue that it is critical for the physicists now flirting with molecular biology to sit down with their new colleagues and agree on what are the important questions. "The word 'function' doesn't exist in physics, but physicists are going to have to learn about it," says Hopfield. "Otherwise they will be off playing in a sandbox by themselves."

Onuchic believes that immersing young physicists in the culture of biology is the key. At the CTBP, postdocs train in both disciplines simultaneously, developing projects that involve two labs, one in biology and one in physics. They attend two sets of group meetings, and so learn the language and mentality of both disciplines at the same time. "They get inside the culture of the two fields," Onuchic says. "They get comfortable with the vocabulary and the journals. Life in both labs is more important than any classes you can take."

Time will tell whether the new generation of biological physicists avoid becoming the lonely children of biology. But for now, the prospects look bright. "We have always been the odd kids in the playground," says Onuchic. "But we never gave up, and now we are becoming very popular."

Jonathan Knight is Nature's contributing correspondent in San Francisco; additional reporting from Natasha McDowell.

1. Hopfield, J. J. *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
2. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. *Nature* **402**, Suppl. C47–C52 (1999).
3. Rappel, W. J., Thomas, P. J., Levine, H. & Loomis, W. F. *Biophys. J.* **83**, 1361–1367 (2002).
4. Barkai, N. & Leibler, S. *Nature* **403**, 267–268 (2000).
5. Frauenfelder, H., Wolynes, P. G. & Austin, R. H. *Rev. Mod. Phys.* **71**, S419–S430 (1999).

Opportunities in Biology for Physicists

♦ www.aps.org/meet/biology-physics/index.html

The many dangers of relying on a DNA database

It may be convenient for the police, but the use of DNA as evidence is full of pitfalls.

Sir — Everyone's DNA profile should be included in a database, according to Robert Williamson and Rony Duncan in their Commentary "DNA testing for all" (ref 1.). This approach has much to recommend it to the police and the prosecution, but it will do little to increase the accuracy of any prosecution in terms of justice.

In Correspondence², David Ehrenfeld invokes a 1907 detective story to comment on problems of classical fingerprinting still relevant today³. If such problems are still arising after nearly 100 years, how much more likely are problems with a universal DNA-profile database only a few years after the introduction of this technology?

One such problem concerns depositing DNA by touch when items are handled. It is possible to detect such handling successfully⁴, a fact already used in a criminal case in Canada⁵. Who can remember every item that he or she has touched for longer than one minute in the past week? When the knife you used in a restaurant is identified as a murder weapon, what is going to be your defence?

There are other major problems associated with a universal DNA-profile database. Would foreign visitors be profiled on entry? What happens if the person who matches the crime-scene DNA

profile has an alibi at a distant place? The database proposal places a reverse onus on accused people to show their innocence, even when there is no other evidence to link them to the crime. What are the error rates for each step in the DNA-profiling procedure? Because the probability of a duplicate match is so low, bayesian theory suggests that when there is no other evidence, error is likely to be a higher probability than a correct match.

Although the safeguards Williamson and Duncan suggest are implementable in the United Kingdom, Australia and other industrialized countries, they are impossible in countries such as South Africa because the cost of independent forensic laboratories is prohibitive. In many countries, most sample collection is done by police officers, not scene-of-crime scientists, and a lack of resources makes independent testing impossible. Samples from suspects, even in the United Kingdom, are taken at police stations. Would independent DNA-testing stations be set up?

Finally, of course, computerized DNA databases are no more immune to hacking than any other database.

The criminal-justice system is weighted against the defendant in many countries.

The defendant often uses a legal-aid lawyer with little or no knowledge of DNA profiling; he or she is in front of a judge with little or no experience of DNA evidence; the defence is only informed of any DNA evidence a couple of days before the trial starts, resulting in little or no chance to verify independently data presented to the court, let alone reprofile the sample.

I accept without reservation that DNA evidence is a very useful tool in criminal prosecutions. However, the police and forensic services in the United Kingdom have at times failed to obey the rules with regard to DNA evidence (see, for example, cases referred to the Court of Appeal Criminal Division, refs 6 and 7). Even now, the system has faults that need to be addressed before any future developments are considered.

Ralph Kirby

Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, PO Box 94, Grahamstown, South Africa 6140

1. Williamson, R. & Duncan, R. *Nature* **418**, 585–586 (2002).
2. Ehrenfeld, D. *Nature* **418**, 583 (2002).
3. Cyranoski, D. *Nature* **417**, 676 (2002).
4. van Oorschot, R. A. H. & Jones, M. K. *Nature* **387**, 767 (1997).
5. R. v. Xie [2000] ABOQB 478.
6. R. v. B [2000] EWCA Crim. 42.
7. R. v. Weir [2000] EWCA Crim. 43.

Could we trust every future government?

Sir — R. Williamson and R. Duncan's inference in their Commentary recommending DNA testing (*Nature* **418**, 585–586; 2002), that governments can be trusted throughout people's lifetimes, is demonstrably false. Their proposed independent DNA database and laboratories could be safeguarded against many, perhaps all, misuses, but could never be protected from coercion by the state.

In the recent past, homosexuality was illegal in the United Kingdom and being a Jew was illegal in Germany. Neither of these attributes is deducible from someone's DNA, of course, but it is not outside the realms of possibility that, in future, traits that have been declared 'illegal' could be detectable in this way. Names and addresses of people falling into some 'illegal' category would then be instantly accessible.

Freedom is valuable because it has taken many generations to build, yet it can be demolished in the time it takes to fight a

war or the single day necessary to hold an election. A state cannot be trusted to host independent records of people's DNA for their entire lifetimes, particularly if we cannot remember abuses against freedom that happened within living memory.

Adrian Bowyer

Department of Mechanical Engineering, University of Bath, Bath BA2 7AY, UK

Planted 'evidence' weakens case for DNA

Sir — R. Williamson and R. Duncan in their Commentary (*Nature* **418**, 585–586; 2002), recommend universal testing "as a deterrent from crime for all members of the community [that] would make the task of catching criminals easier for police".

The obvious flaw in this proposal is the ease with which biological material can be moved from one place to another. Anyone with criminal intent could simply collect biological material (cells from a toothbrush, perhaps?) and plant the evidence at the scene of their crime. In the society proposed by Williamson and

Duncan, respect for DNA evidence would be total, and thus innocent people would more easily be convicted.

Oliver Flint

Bristol-Myers Squibb, PO Box 5400, HW17-2.04, Princeton, New Jersey 08543, USA

Free consanguinity testing for all

Sir — In their recent Commentary (*Nature* **418**, 585–586; 2002), R. Williamson and R. Duncan discuss the use of universal neonatal DNA screening for forensic purposes. Such an approach would take a huge investment for at least 20–30 years before showing any benefit in terms of crime prevention. It also seems bizarre to treat each individual as likely to commit a serious crime from the time of birth!

Free DNA testing for all is essential, but for very different reasons. In most cultures, at least a small percentage of children are born through infidelity. Added to that, as many as 10% of couples now use assisted reproduction, often involving the use of gametes from unknown donors. Children

born through such practices are likely to be of similar age and to grow up in the same area. A significant percentage of couples may, unknowingly, be closely related.

These serious problems have not received sufficient attention, and are likely to be contributing to the birth of children with serious genetic conditions. It is therefore essential that the availability of wider reproductive choices in modern society should be coupled to education about the risks of consanguinity and the provision of free DNA testing for all people of reproductive age, competently administered by an independent authority under strict regulations. This would empower every couple to make an informed choice, based on their genetic relatedness, on whether to proceed with childbearing or to seek alternative solutions. Such a service will result in immediate benefits for both the participating individuals and society at large.

The information acquired through consanguinity testing should be archived by the testing authority to facilitate long-term follow-up. The consanguinity database could be made accessible to the police authorities searching for matches with samples taken from crime scenes, but the identity of any positive samples should be revealed to the police authorities only after proper, independent authorization.

Consanguinity testing can thus complement DNA profiling of convicted individuals by police authorities, without the negative connotations of neonatal DNA screening.

Panos Ioannou

*Cell and Gene Therapy Research Group,
Murdoch Children's Research Institute,
Royal Children's Hospital, Flemington Road,
Parkville, Victoria 3052, Australia*

Why response-mode research loses out

Sir—Knowledge inferred from DNA sequences can make an important contribution to taxonomy, as discussed in Correspondence by Diethard Tautz *et al.* (*Nature* **418**, 479; 2002). But I fear that structural research-funding problems stand in the way of initiatives such as this, at least in the United Kingdom.

No charitable foundation is dedicated to the support of work in non-medical evolution and taxonomy; the UK's Natural Environment Research Council (NERC) is the main funding agency. However, as witnessed by my drawerful of alpha-rated but unfunded proposals, it does not appear to have enough money to go around. Systematic sequencing is a long-haul activity that needs long-term, modest

funding. But proposals for the continuation of any line of work have to compete with the entire range of what is newest and best in the field (in my case, the whole of marine science), a competition that no proposal to do “more of the same” can possibly win, no matter how rare or how choice the samples.

Worse still, despite all the evidence that centralization leads to waste and failure, NERC sequesters a large proportion of its funds in thematic programmes, which many believe distort the research enterprise and give preferential funding access to largely self-selected coteries. Two questions that cannot be answered are these: “How many funded thematic projects would have succeeded in open competition?” and “Do thematic programmes provide better science than the (many more) response-mode proposals that would otherwise have been fundable?”. History provides the best guide to an answer. Before research councils were invented and before they started controlling research direction, British science was internationally well-regarded and successful. Is it still? How about a decade or two of response-mode-only funding, with all proposals in equal and open competition?

Bernard L. Cohen

*Institute of Biomedical and Life Sciences, Division
of Molecular Genetics, University of Glasgow,
Pontecorvo Building, Glasgow G11 6NU, UK*

Europe is not yet ready for a research council

Sir—The suggestion of a European Research Council (ERC) is again in the news (*Nature* **419**, 108–109, 2000; *Nature* **418**, 259, 2002). In the context of continued underfunding of basic research in many European Union (EU) countries, with the possible exception of the United Kingdom, and the limitations of EU Framework funding, an alternative source of funds for competitive grants has to be welcomed. However, there are many problems to be anticipated if the ERC is to be effective.

Setting up an ERC now would be premature. First, the chronic state of underachievement by most EU countries in the basic sciences desperately needs attention. Governments and national funding agencies must spend substantially more money, and seriously promote more democratic and satisfying career structures, permitting scientists to fulfil their potential. Failing this, we shall never be able to compete at the European level with the United States. There seems little sign that EU member states realize the seriousness of the situation.

In principle, an ERC with an adequate level of funding and a mechanism for administering funds could at some stage make European science more competitive. But an ERC with a life-science budget less than that of the NIH, or similar to that of the UK's Wellcome Trust, for example, will be inadequate and counterproductive. Equally, an ERC that is subject to the constraints of the over-bureaucratic Framework programmes would be a disappointment. To be effective, an ERC would also need to have the means, the will, the power and the independence to change attitudes. It would need to set standards of training, career development and salaries for young scientists in much the same way as the Wellcome Trust over the past 15–20 years has dramatically raised standards in the United Kingdom.

It seems inconceivable that EU member states, many of whom underfund and fail to democratize their own national research bases, will sufficiently fund a supra-national agency which they could perceive as undercutting their control over scientific policy. Any significant input of funds from the current Framework programmes, already inadequate in relation to demand, also seems unrealistic.

If insignificant amounts of money are made available to an ERC, why bother with it? It would be just another peripheral funding body, allowing member states to shrug off the more serious problems at home. In the life sciences, member governments could more effectively simply increase their contributions to the Human Frontier Science programme (well administered but chronically short of funds), or to the European Molecular Biology Organisation (EMBO) to allow expansion of its excellent graduate-student and postdoctoral programmes.

I. Barry Holland

*Institut de Génétique et Microbiologie,
UMR-CNRS 8621, Bat. 409, Université Paris XI,
91405 Orsay cedex, France*

See also the Commentary, pages 249–250 of this issue — Editor, Correspondence

Element of confusion

Sir—I enjoyed your News Feature “A very brief encounter” (*Nature* **418**, 815–816; 2002) on the very heavy elements and their reactivity. I was a little perplexed, however, by the periodic table on page 816 that identified element 110 as H, particularly as H is already included elsewhere in the table (correctly) as element 1, hydrogen. Element 110, of course, is not yet named.

Richard Joyner

*Department of Chemistry and Physics,
Nottingham Trent University, Burton Street,
Nottingham NG1 4BU, UK*

A fresh start for European science

The scientific community must take up the challenges set by EU objectives.

Wilhelm Krull

Europe is going through rapid changes, and it lacks something it urgently needs in almost all areas of research policy-making: a strong Europe-wide voice for science and scholarship. Even more pressing for the European Union (EU) is the lack of a central, financially powerful funding institution, able to make decisions based on qualitative judgements by the best researchers worldwide.

Apart from a few research areas such as astrophysics, space research, nuclear physics and, to a limited extent, molecular biology, Europe suffers from an almost total lack of transnational support of basic and strategic research. European research needs institutional reforms at all levels to keep pace with the changes inherent in becoming a knowledge-based economy. It is increasingly desirable, even urgent, to establish pan-European funding structures that can create both a cooperative climate for the development of new ideas, and an institutional environment that will encourage competition among Europe's best researchers to produce more cutting-edge results.

Catalytic activity

In this wider context, a new European research council could be a spearhead of institutional reform, a catalyst of new inter- and transdisciplinary research activities, a creator of new transnational funding opportunities for young researchers and, last but not least, a provider of the research-friendly administrative and organizational



environment that is urgently needed to attract foreign researchers (see *Nature* **419**, 108–109; 2002).

The Lisbon declaration of March 2000 set the political goal of developing the EU into “the most competitive knowledge-based economy in the world by 2010”. At a meeting in Barcelona, also in March 2000, the EU Council agreed to increase investments in research and development (R&D) across the EU from today's 2% of gross domestic prod-

uct to 3% by 2010. These initiatives are not meant to become yet another manifestation of politicians' rhetoric. It is time for us in the scientific community to respond seriously to the challenges that they provide.

An immediate challenge

The recently completed first round of EU benchmarking exercises clearly demonstrates that there is neither time nor room for complacency. The EU is lagging behind the United States not only in R&D expenditure and overall output, but also on such widely accepted indicators of scientific and technological productivity as publications per inhabitant and citations per scientific publication (see Figure, left), as well as on several patent indicators. Even more striking are the huge differences in scientific and technological productivity between the EU member states, with the Nordic countries clearly taking the lead — as they did in the recent Programme for International Student Assessment (PISA) study of schoolchildren's mathematics and literacy in 32 countries around the world (see www.pisa.oecd.org). Some large countries, such as Germany, achieved only slightly above-average results, which is disappointing.

Publications and patents are intermediate outputs of research, so they are only a partial measure of the achievement of broader goals in the advance of knowledge. Nevertheless, they can serve as starting points for taking a



Lagging behind: Europe's citation rate remains below that of the United States.

closer look at the configuration of research systems and innovation systems. Areas of particular concern in Europe include the distribution of excellence among research institutions, the infrastructure and other base conditions, and funding opportunities for the most talented researchers. We need to develop attractive career structures which enable those researchers to pursue their own ideas much earlier and more independently than in most current systems. The continual flow of highly qualified researchers between countries and between the public and private sectors also requires more flexibility and permeability. Qualifications gained in national institutions must be valid throughout Europe.

In particular, the publicly financed universities, research institutes and funding bodies in Europe have an important role in providing the next generation of researchers. Our society's well-being depends on the quality of these people's talents and their ability to meet the challenges of the future. Healthy universities provide the basis for a globally competitive R&D workforce. But it is only in the framework of international research communities that a university system, research organization or funding institution can attain the highest standards of achievement.

Open approach

Although research organizations have begun to invite foreign independent experts to take part in their assessment exercises, most research councils are still overwhelmingly nationally oriented when it comes to peer review of proposals. They rely almost completely on the ability of colleagues in the same disciplines at institutions in their own countries to select high-quality projects for support. Even if one believes that this system works well for most subjects, it is easy to identify areas where mediocre researchers endorse each other and nepotism is rife. These self-congratulatory circles need to be broken, by denationalizing grant-making and peer review throughout Europe.

Compensating for weaknesses in research training schemes and overcoming deficiencies in national councils' assessment processes are but two of the many pressing reasons for rethinking European research funding structures. There are other equally important reasons for establishing a research council at the European level (see Box).

Of course, these broadly defined areas of activity need to be specified in more detail, but they are a first set of indications of funding needs and priorities that are not adequately covered by national research councils or agencies, nor by the European Commission and its multifaceted (though mainly precompetitive) industry-oriented R&D programmes. These developments cannot just be added to existing funding structures — they require different

Council activity

Possible functions of a European research council:

- Create medium- to large-scale facilities beyond the scope of one country (for example, clustering of imaging techniques, establishing central animal repositories), and make better use of existing large facilities.
- Set priorities in transdisciplinary research and provide incentives for picking up new areas of innovative science and scholarship.
- Add a clear European dimension to the competition for some of the most prestigious grants and awards.
- Establish leading-edge collaborative centres of appropriate size in basic and strategic research areas that call for integrative approaches from different disciplines.
- Offer additional funding opportunities and new career structures for young postdoctoral researchers, and thus enable them to pursue their own ideas in an internationally supported, highly stimulating environment.

ways of strategic thinking and of interacting within the discipline concerned and more widely. They also need a more research-friendly, less bureaucratic but well-organized and rigorous selection process, which cannot obviously be grafted onto any existing mechanism.

Several existing pan-European research organizations, such as the European Molecular Biology Organization (EMBO) or the European Science Foundation (ESF), could provide some of the functions listed in the Box. But none seems willing or able to cope with the huge challenges of becoming the main funder of scientific and scholarly excellence in Europe.

To put it another way, at the European level we have too many weak organizations. To develop them into well-equipped, financially powerful institutions with executive authority may not be impossible. But because of their current stakeholders' limited perspectives, reform could be a cumbersome process with meagre results. It may suffice to mention one such attempt in the early to mid-1990s: the reappraisal of the



ESF's mission. Despite clearly identified needs, a lot of effort and long debates, the result is at best an enlarged pan-European mission statement for the ESF. Because neither its member organizations nor the European Commission gave it enough extra funds, the ESF has more or less remained what it always was for heads of national research councils: a little item at the bottom of the agenda.

To secure a framework of autonomous decision-making for a new funding body will not be straightforward. It will take a lot of courage and openness by the heads of the primarily nationally oriented research organizations, as well as by the EU member states' governments, and possibly also beyond the EU's current borders. They will need to grant the utmost degree of freedom and, at the same time, a great deal of money to an institution that will have to prove that it can live up to its aspirations. But it is certainly worth the effort, if we Europeans

want to get a little closer to becoming the world's most competitive knowledge-based economy by 2010.

Given the current need to reconfigure the political architecture of the EU, there is now a window of opportunity for rethinking and realigning our research funding structures. This will enable us to move further away from the old EEC

approach, which was forced by legal constraints to demand an industry-oriented justification for short- to medium-term R&D programmes at the European level, while at the same time paying tribute to '*juste retour*' — the principle that researchers in any country receive roughly the same percentage of funds as their government contributed to the pot.

This has led us into the rather strange situation where regionally oriented innovation policies are dealt with at the European level, while genuinely internationally oriented basic and strategic research activities find only very limited support beyond national borders. It is time for a change. Now that today's politicians seem to understand the necessity of taking a long-term view of capacity building in science and scholarship, it may no longer be a frustrating, Sisyphus-like exercise to strive for a thorough restructuring of research funding at the European level. Let's give the bold vision for a European research council a chance! ■

Wilhelm Krull is the secretary general of the Volkswagen Foundation

(www.volkswagenstiftung.de), Germany's largest private funder of higher education and research, at Kastanienallee 35, 30519 Hannover, Germany. He is a founding member of Euroscience (www.euroscience.org).

Qualifications gained in national institutions need to be valid throughout Europe.

Nurturing a view of human nature

Our genes ensure that our minds are never 'blank slates'.

The Blank Slate: The Modern Denial of Human Nature

by Steven Pinker

Allen Lane (Penguin)/Viking; 2002. 561 pp.
£25/\$27.95

David L. Hull

Steven Pinker begins the preface of his book with the exclamation: "Not *another* book on nature and nurture!" In fact, the contrast between nature and nurture is only one of the topics that Pinker addresses in this book, but throughout his exposition I shared the dismay of his hypothetical reader. Few people nowadays who know anything about genetics, evolution and development can maintain that the human mind is a blank slate. Every phenotypic trait exhibited by living creatures is the result of a complex interplay between nature and nurture, genes and environment. No characteristic is entirely innate or strictly acquired. Even the seventeenth-century philosopher John Locke, who coined the phrase *tabula rasa* or 'clean slate', acknowledged that the mind has "innate capacities". With respect to minds being blank slates at birth, surely Pinker is beating a straw man with a dead horse.

In most of his book, Pinker accepts the nature–nurture distinction and complains that his opponents do not give nature sufficient weight, but towards the end he stumbles on quite a different conclusion. Perhaps the nature–nurture distinction itself is at fault. Pinker acknowledges that the human genome cannot possibly specify every connection among our neurons. The environment, in the sense of "information encoded by the sense organs", also plays a role, but even these two causal factors are not enough. Chance also contributes to development: for example, one twin lies one way in the womb, the other lies a different way. The non-genetic component of personality may well be to some extent the outcome of "neurodevelopmental roulette", as Pinker puts it. "Just as the 'genetic' term in the behavioural geneticist's equation is not necessarily genetic, the 'environmental' term is not necessarily environmental," he explains. In short, distinguishing between genes and environment may not be a very perspicuous way of dividing up the developmental pie. Support for this conclusion comes from the difficulty of saying anything both clear and straightforward using this distinction.

Pinker called his book *The Blank Slate*. He thinks that lots of people believe that we are born with minds that are totally blank; I do not think that this view is so widely accepted. The book's subtitle is *The Modern*



Slated: Steven Pinker says the mind is not like a clean sheet of paper to be written on by experience.

Denial of Human Nature, and here again I disagree with him about how widely distributed this rejection is. Pinker thinks that a large proportion of the reading public denies the existence of human nature. I have taught philosophy to hundreds of undergraduates over the years, and they all claim to believe in human nature. They disagree widely on what this nature is, but like Pinker they take for granted that it exists. All humans, and only humans, exhibit a particular characteristic. Anyone who lacks this characteristic is abnormal, not truly human.

From the perspective of embryological development, labelling certain traits and genes 'abnormal' makes some sense. Certain alleles perform a particular function better than other alleles. For example, those people who suffer from sickle-cell anaemia possess an abnormal gene. So do those of us with blue eyes. Both traits are due to damaged genes. It may be appropriate to consider certain traits and genes as being abnormal from the perspective of embryological development, but it makes no sense with respect to evolution. Every gene that we now possess was once unique. Natural selection is the primary mechanism for changing gene frequencies. As a gene decreases in frequency, it does not become increasingly abnormal.

Pinker terms himself an evolutionary psychologist, but one of the requirements of evolution by means of natural selection is

variation, both genotypic and phenotypic. Variation is the essence of selection, and Pinker knows it. He estimates that roughly half of the variation in traits is due to genetic variation, and that half of the human genome that functions in development is involved in producing the human brain. Yet, at the level of human minds, none of this variation matters. Way down deep, all minds are essentially the same.

Of course, the easy response is that all this variability can be dismissed as abnormal. Pinker says that his book is primarily about human nature — an endowment that is universal in 'healthy' members of *Homo sapiens*. In an appendix he presents Donald Brown's list of almost 400 human universals, one of them being to distinguish between normal and abnormal states. But Pinker admits that many of these traits are universal only among 'normal' people or in 'normal' circumstances. However, as much as these superficial universals vary, at a deeper level he thinks that there are truly universal universals; that beneath superficial differences, our species has a "psychological unity".

What puzzles me is why Pinker adopts as one of the essential tenets of evolutionary psychology a position that runs so counter to evolutionary theory. Why have evolutionary psychologists saddled themselves with the monomorphic mind? We do not have monomorphic blood types, monomorphic eye

GETTY IMAGES

Reptiles of the republic

The republic of Costa Rica is home to almost 400 known species of amphibian and reptile. In *The Amphibians and Reptiles of Costa Rica: A Herpetofauna Between Two Continents, Between Two Seas* (University of Chicago Press, £75), Jay M. Savage provides an in-depth guide to the biology of each species, with details of their anatomy, behaviour, systematics and distribution, as well as an identification guide. The book also contains many splendid photographs by Michael and Patricia Fogden, including this striking image of the snake *Bothriechis schlegelii*.



colour, or monomorphic hearts. So why are evolutionary psychologists so insistent that we all have monomorphic minds? One possible answer is to avoid the charge of racism.

Those people, whoever they may be, who actually believe that all minds are blank slates are off the hook with respect to being called racists; after all, blank is blank. But evolutionary psychologists insist that genes play an important role in making us what we are, and it is possible that these genes cluster. Some subgroup of human beings might have more of the 'good' genes than other subgroups. As a result, a charge of racism might have some justification. But if all the variation that occurs among people is only surface variation, then evolutionary psychologists are also off the hook: we all have fundamentally the same mind. Those evolutionary biologists who think that both genes and traits are distributed in highly complex ways can also avoid being termed racists, but their argument is a good deal more complex.

Much of this discussion presupposes that in order to have the same rights, we must all be fundamentally the same. Pinker argues that, at a very fundamental level, all human beings are the same. Hence, we all have the same rights. But he also quotes Ernst Mayr as arguing for a much more sophisticated position: that equality does not require identity. Why can't we all have the same rights even if in many ways we are essentially different? That is the question.

I oppose the blank-slate view of human minds as strongly as Pinker does. I just do not think that it is as widespread as he does. I also think, contrary to Pinker, that a belief in the existence of something properly

termed 'human nature' is very widespread. The trouble is that these 'natures' are highly variable, as indeed they must be if they are to evolve. They are innate but neither fixed nor universal. In *The Blank Slate*, Pinker presents an overarching view of the world in a way that quite a few readers will find seductive. The preceding objections aside, I find myself in basic agreement with Pinker's world-view. That I can accept so much of what he has to say while rejecting one of the positions that he takes to be fundamental implies that possibly this position is not really so fundamental. ■

David Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60208-1315, USA.

Dog-days at the data factory

Pavlov's Physiology Factory: Experiment, Interpretation, Laboratory Enterprise
by Daniel P. Todes
Johns Hopkins University Press: 2002.
512 pp. \$58

Steve Sturdy

Ivan Pavlov is remembered chiefly for his work on the conditioned reflex, one of the fundamental concepts of modern physiological psychology. But when in 1904 he became the first physiologist to receive a Nobel prize, he had barely started to investigate conditioning. Rather, the prize was awarded in recognition of his ground-

breaking research into the nervous integration and control of digestive secretion. It is this earlier work that Daniel Todes examines in *Pavlov's Physiology Factory*.

Pavlov's rise to international eminence effectively began in 1891, when he was appointed head of the physiology division of the sumptuous new Institute of Experimental Medicine in St Petersburg, Russia. Physiology at that time was in transition from being a field of largely individual endeavour to one of increasingly collaborative effort. Pavlov stands out for the sheer degree of managerial control that he exerted over the work conducted in his laboratory. In part this was due to the nature of his workforce. Most of his co-workers were medical students spending a brief sojourn in the laboratory en route to a career in medical practice. By specifying the topics they addressed and the precise methods they used, Pavlov ensured that their efforts were channelled to produce meaningful data for his own research programme.

But Pavlov's organizational innovations were also inspired by developments elsewhere in Russian society. His enthusiasm for laboratory science was part of a more general commitment to social modernization, which idealized the factory system as the most effective means of harnessing the productive capacity of workers. The managerial regime that he imposed on his laboratory was a self-conscious adaptation of this ideal for the purpose of producing scientific knowledge. And as in the factory, technology played an important role in determining the activities of semi-skilled workers.

Pavlov's investigations revolved around several surgically modified but otherwise healthy dogs — 'chronic' as opposed to 'acute' vivisections — in which a succession of researchers could observe the passage of food and the flow of secretions through the digestive tract. It was the development of these "dog technologies", as Todes calls them, as much as Pavlov's intellectual supervision, that made possible the continuous production not just of physiological facts and theories, but also of such commercially successful therapeutic commodities as natural gastric juices.

Literary products such as theses and publications also issued from Pavlov's lab, and he maintained strict control over all of them. Final responsibility for the theoretical elaboration of experimental data remained firmly in Pavlov's hands. Here, too, he found inspiration in the ideal of the factory, although this time of a metaphorical kind. He saw the alimentary system as a complex chemical factory in which the work of the various digestive organs was regulated by nervous reflexes to ensure the most efficient digestion of different foodstuffs.

Todes makes clear how Pavlov's scientific conclusions were shaped by such presuppo-



Production line: for his early work on digestion, Ivan Pavlov collected dogs' gastric secretions.

sitions. Pavlov was acutely aware that his results often varied widely, not just from one animal to another but within the space of a single experiment, in response to a host of often unidentifiable disturbances in the experimental system. Indeed, his later work on conditioned reflexes grew out of his efforts to characterize the role of "psychic" phenomena in producing such perturbations. But it was precisely this variability that enabled Pavlov to select those data that best confirmed his supposition that digestive processes were purposively and selectively regulated, and to explain away any results that contravened the supposition. Pavlov's wishful thinking eventually became apparent, however, particularly when the demonstration by W. M. Bayliss and E. H. Starling of the humoral control of pancreatic secretion raised doubts about his claim to have identified more selective forms of nervous regulation.

Such revelations did not deter the Nobel committee, however, who by that time were considering Pavlov for the prize in physiology or medicine. Initially, his supporters on the committee had cited the theoretical significance of his conclusions as sufficient grounds for awarding him the prize. But as the validity of those conclusions began to be challenged, they instead invoked Pavlov's methodological innovations and his demonstration of the promise of concerted laboratory research for the pursuit of medical knowledge. Pavlov's elevation to the pantheon of Nobel laureates thus rested, in the final analysis, as much on the way that he made manifest the new values and aspirations of collective scientific endeavour as on any specific contribution to scientific knowledge.

Pavlov's achievements are in no way diminished by such scrutiny. On the contrary, by revealing the technical, intellectual, managerial and literary skills that led to the

success of Pavlov's endeavour, Todes broadens our appreciation of the far-reaching contribution that Pavlov made to the development of modern science. Todes has achieved an impressive feat of scholarship, combining meticulous research with analytical clarity, which does full justice to his compelling subject. ■

Steve Sturdy is in the Science Studies Unit, University of Edinburgh, Edinburgh EH8 9LN, UK.

A stroll with the moulds

Mr. Bloomfield's Orchard: The Mysterious World of Mushrooms, Molds, and Mycologists

by Nicholas P. Money

Oxford University Press: 2002. 224 pp. \$26

Elio Schaechter

Countless stories can be told about any large group of living things, and fungi are no exception. Fungi are quite familiar to us: we enlist their help in making bread, wine and beer, and are vexed by the disease that they cause in humans, animals and plants. We know them as unicellular yeasts, filamentous moulds and complex-looking mushrooms. Less widely appreciated is their most important pursuit: they are the essential decomposers of vegetable matter. Quite simply, life on Earth would not be possible without the recycling activity of fungi.

Fungi have a wide repertoire of life cycles, a large variety of shapes and structures, and a strong propensity to interact with other organisms in manners benevolent and otherwise. Quite a few of the stories about each of these properties can be found in



Red and green: fungi such as the scarlet elf cup aid nutrient cycling by breaking down vegetation.

Mr. Bloomfield's Orchard. The main topics are the life cycle of moulds and water moulds (oomycetes), aspects of fungal biomechanics, and fungal parasitism of plants and humans.

Nicholas Money gives considerable attention to subjects in his own research area, such as cytoplasmic turgor and the electrical and mechanical properties of hyphae. These lead to some lively examples: stinkhorns that make geodesic domes (*Clathrus*), hyphal filaments that can pierce bullet-proof vests (let alone human skin), spores that trap air bubbles for buoyancy, and moulds that quarry rocks. He also describes in some detail the ingenious strategies for spore dispersal in mushrooms, plant pathogenic moulds and water moulds, and helpfully explains stories that tend to be quite intricate. For instance, the life cycle of certain rust-causing fungi includes four kinds of spore, two separate plant hosts, and three acts of plant penetration.

The contributions and quirks of several notable mycologists also get a mention. In the early 1900s, Reginald Buller did much to explain how fungi, especially mushrooms, disperse their spores. It is now believed that the mechanism involves the formation of a droplet at the base of the spore. By a mechanism that involves changes in surface tension, this 'Buller's drop' helps to catapult the spore away from its native gill with astounding acceleration. Buller, an Englishman transplanted to Winnipeg in Canada, had a distinctive personality. Needing to be dark-adapted to make his observations, Buller used horse blinkers to keep out the light as he walked through the sunshine from his hotel to the university. Money also portrays another colourful mycologist, Curtis Gates Lloyd, who circumvented the risks of peer review by publishing his own journals.

Books on fungi written for the non-specialist are few in number. Pre-eminent in this genre is E. C. Large's *Advance of the Fungi* (Jonathan Cape, 1940). In this much-revered book, Large focused on plant pathogens and described their activities in a lively and informative style that infused the reader with enthusiasm for the subject. Money's book follows this tradition. His writing is accommodating and personal, with occasional chummy asides. He makes no bones about his enthusiasm for the fancy shapes and forms of moulds. He says of yeast: "Only a biochemist can be satisfied with such dull architecture when the grandeur of other fungi can surpass the Palace of Versailles."

Whether one agrees with this view or not, one can enjoy the book for the choice of topics and the clarity of the writing. It can be recommended to all nature lovers, regardless of background, who want to know more about fungi. ■

Elio Schaechter is in the Department of Biology, San Diego State University, San Diego, California 92182, USA.

Animal reflections

Marc Bekoff

Researchers are interested in animal awareness because they are curious to discover what animals might know about themselves. There are, however, long-held and polarized views about the degree of self-awareness in animals. Some people believe that only great apes have 'rich' notions of self — knowing who they are and/or having a 'theory of mind', which means being able to infer the states of minds of others — whereas others argue that it is methodologically too difficult to address this question because animal (like human) minds are subjective and private. Many in this latter category do not attribute any sense of self to animals other than humans, and some, dismissing behavioural and neurobiological research on animal cognition, wonder whether animals are conscious of anything at all.

What might animals know about themselves? Most studies of animal self-awareness have been narrowly paradigm-driven. The 'red spot' technique was first used by Gordon Gallup to study animal self-awareness in chimpanzees; it and variations have been used on great apes and monkeys, as well as on a few dolphins and elephants. For primates, a spot is placed on the forehead of an anaesthetized individual and self-directed movements towards the spot are scored after he or she awakens and catches sight of themselves in a mirror, a high score indicating the presence of some degree of self-awareness. But in some cases, the data are derived from tests on small numbers of individuals, many of whom fail it because they do not make self-directed movements towards the spot. Those who pass the test might not be representative of wild relatives because they have had extensive

human contact and previous experience with mirrors, factors that might influence their trainability and willingness to use a mirror. Those who fail the test might show some sense of 'self' in other contexts, and other individual differences might also play a role.

Laboratory experiments can overlook the fact that many animals have evolved in complex webs characterized by variable social, sensory and other environmental features. A single technique based solely on visual cues is not the only valid test of self-awareness because many animals regularly use sensory modalities other than vision, either singly or together. Most wild animals do not know what they look like — although birds can acquire information about their plumage colour from reflections in drinking water — but they seem to know what they sound or smell like. Many birds distinguish their own from others' songs and numerous mammals discriminate their own from others' scents. Some rodents use odour to distinguish 'self' from 'other'; they preferentially associate or play with those who smell like them, or shun individuals who smell like them to avoid inbreeding.

Discriminating between 'own' and 'other' is not necessarily an indication of self-awareness, but in most cases we don't know if it is or isn't. Although no one has suggested that birds or rodents are self-aware to the degree that chimpanzees might be, we do not know precisely what these animals know about themselves. We must be careful neither to imbue animals with unknown cognitive capacities nor to rob them of skills that they might possess.

Researchers often dismiss non-primates when they investigate self-awareness. Social carnivores, such as grey wolves, who live in closely knit, orderly and efficient packs, could turn out to have well-developed self-awareness. Wolves engage in coordinated hunts and rear their young communally. Analyses of the behaviour patterns of individual wolves during these (and other) activities show that they have at least some awareness of what they are doing, what others are doing, and where each is located spatially. Similar observations have been compiled for other species.

Comparative data suggest that if we rely solely on mirror studies, we have been looking for self in the wrong places and on the wrong faces. The evidence also indicates that no other animal has the same degree of self as that possessed by most humans (an 'I-ness' or 'I-self', perhaps mediated by language). On the basis of traditional measures, many animals appear to have a sense of 'body-ness' or 'mine-ness', but not a sense of 'I-ness.'

Answers to challenging questions about self-awareness have wide-ranging conse-

Awareness

On traditional measures, many animals appear to have a sense of 'body-ness' or 'mine-ness', but no sense of 'I-ness'.

quences because they are often used by researchers and lawyers as a litmus test for defending the sorts of treatments to which animals can be ethically subjected. However, it is not clear that self-awareness or other cognitive capacities should be used for such decisions. Some argue that a sense of 'I-ness' is morally relevant and necessary for experiencing pain. But even if an animal does not know 'who' she is, this does not mean she cannot feel that 'something painful is happening to this body'. Even though the experience of pain might not be the same across species, individuals of different species can still suffer their own type or version of pain. Self-awareness is not a reliable test for an objective assessment of well-being.

The concept of animal self-awareness remains open to different interpretations, but we will probably learn more about the mysteries of 'self' and 'body-ness' by using non-invasive neuroimaging techniques in combination with cognitive ethological studies. If we look at 'self-awareness' as 'body-awareness', we might also discover more about how animals think and the perceptual and neurobiological processes underlying various cognitive capacities. Darwin's ideas about evolutionary continuity, together with empirical data ('science sense') and common sense, caution against the unyielding claim that humans — and perhaps other great apes and cetaceans — are the only species in which some sense of self has evolved.

Marc Bekoff is in the Department of Environmental, Population, and Organismic Biology, University of Colorado, Boulder, Colorado 80309-0334, USA.

FURTHER READING

Bekoff, M. *Minding Animals: Awareness, Emotions, and Heart* (Oxford Univ. Press, New York & London, 2002).

Bekoff, M., Allen, C. & Burghardt, G. M. (eds) *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition* (MIT Press, Cambridge, Massachusetts, 2002); see especially essays on self-awareness by Gallup, G. G., Anderson, J. R. & Shillito, D. J.; Mitchell, R. W.; Shumaker, R. W. & Swartz, K. B. Mitchell, R. W. in *Handbook of Self and Identity* (eds Leary, M. R. & Tangney, J.) 567–593 (Guilford, New York, 2002).

Reiss, D. *Nature* **418**, 369–370 (2002).

Rilling, J. K. *et al. Neuron* **35**, 395–405 (2002).

Rilling, J. *et al. Am. J. Phys. Anthropol. Suppl.* **30**, 263 (2000).

Rilling, K. *et al. Biol. Psychiat.* **49**, 146–157 (2001).

L. KENNEDY/CORBIS



Social work: is the pack behaviour of grey wolves indicative of some degree of self-awareness?

Parkfield's unfulfilled promise

Ross S. Stein

The idea that earthquakes are 'time-predictable' underlies many of today's probabilistic forecasts. In a key test on California's San Andreas fault the concept is found wanting, but the news may not be all bad.

How is the Parkfield Earthquake Experiment like technology stocks? They both seemed like a great bet at the time. As the first attempt to track a geological fault to failure, Parkfield (Fig. 1) is without peer anywhere in the world. This 25-km-long stretch of the San Andreas fault has hosted five (possibly six) shocks since 1857, all reaching magnitude 6 on the Richter scale. Not only were these earthquakes roughly the same in size and location, but they were also separated by about the same amount of time, with successive intervals between earthquakes of 24, 20, 21, 12 and 32 years. This pattern of roughly periodic events is bolstered by nearly identical seismograms for the three most recent earthquakes. In addition, the last two shocks, in 1934 and 1966, were both preceded by a magnitude-5 foreshock some 17 minutes before the main shock, several kilometres to the north.

Parkfield's behaviour anchored the concept of time-predictable earthquake recurrence, in which an earthquake strikes when a fault recovers the stress relieved by the previous event¹. This means that the larger the earthquake on a fault, the longer the wait until the next one. With each Parkfield quake similar in size, and a constant loading rate, the next would occur in 22 years — in 1988 — with an uncertainty of about 10 years. And so, in 1986 the United States Geological Survey inaugurated a focused experiment to measure the strain accumulation, capture the nucleation of the next rupture, and watch it propagate². But the earthquake never arrived.

Even though the timing of the Parkfield earthquakes has turned out to be more irregular than envisaged, if the stress accumulated since 1966 has not exceeded the stress released in 1966, the earthquakes would still be time-predictable. But on page 287 of this issue, Murray and Segall³ remove this last refuge. Drawing on 40 years of geodetic measurements, they rigorously estimate the 'seismic moment' of the 1966 event (a measure of the earthquake's size — a product of the fault slip, the area of slip and the elastic stiffness of the crust), and compare it to the accumulating 'moment deficit' (the moment associated with slip that has not occurred since the 1966 quake). Astonishingly, they find that the moment released in 1966 had been fully recovered by 1987. The expected earthquake is thus at least 15 years overdue, approaching a complete cycle.



Figure 1 The San Andreas fault near Cholame, California. Cholame lies at the southern end of the Parkfield region, which has suffered five or six magnitude-6 earthquakes at regular intervals since 1857.

So have Murray and Segall dethroned time-predictability? The authors acknowledge a key caveat: faults are not stressed in isolation. Other earthquakes change the stress acting on the San Andreas fault, which could delay or advance the next Parkfield event. Candidates include the 1983 Coalinga earthquake, 25 km to the east and magnitude 6.5, which might have delayed the Parkfield earthquake by a decade⁴; and a 1993 event⁵, magnitude 4.6 and a few kilometres beneath the focus of the 1966 quake, which might have advanced the next Parkfield earthquake. Although the effects of such stress changes might explain the delay, they cannot rescue time-predictability, because Murray and Segall measure the full moment deficit between 1966 and 1998, which includes the effects of the perturbing shocks. Thus, the moment imbalance remains inescapable.

But does a lone, overdue earthquake

invalidate an earthquake-recurrence hypothesis? Its proponents would argue that it is based on rational physical principles, and that it works better than the Poisson hypothesis, which assumes the timings of earthquakes are random. Further, time-predictability, or at least periodic earthquake occurrence, has been found to occur for some very small 'repeating' shocks at Parkfield and elsewhere⁶. When applied to larger earthquakes, time-predictability works only crudely, and is thus applied with the explicit assumption of large expected variability. What is now needed are tests of time-predictability on other faults. The ingredients of a good test are fast fault-slip rates, moderate-sized earthquakes with short inter-event times, and dense geodetic coverage. All of this points to Japan, where the concept was born.

While the Parkfield experiment can no longer serve as a confirmation of time-

MICHAEL COLLIER

predictability, it is very much alive as a test of slip-predictability, in which the longer the wait, the larger the next earthquake¹. Murray and Segall argue that an earthquake of magnitude 6.6–6.9 would balance the moment deficit that has accumulated since 1966, and the magnitude increases with each passing year⁷. If, when the next Parkfield shock strikes, its magnitude approaches this expectation, the Parkfield experiment would be transformed from a lesson in patience to a prescient success — tracking the nucleation and propagation of a much larger shock with Parkfield's arsenal of instruments would be a bigger scientific prize. And the observing power at Parkfield has just taken a new leap, with the completion of a 2.2-km-deep pilot well with a string of down-hole seismom-

eters, and the proposed 4.0-km borehole that will pierce the fault at seismogenic depths⁸. Perhaps in the end the delay will appear providential, and Parkfield will turn out to be an inspired long-term investment for science. ■

Ross S. Stein is at the United States Geological Survey, Menlo Park, California 94025, USA.
e-mail: rstein@usgs.gov

1. Shimazaki, K. & Nakata, T. *Geophys. Res. Lett.* **7**, 279–282 (1980).
2. Roeloffs, E. *Curr. Sci.* **79**, 1226–1236 (2000).
3. Murray, J. & Segall, P. *Nature* **419**, 287–291 (2002).
4. Toda, S. & Stein, R. S. *J. Geophys. Res.* **107**, 10.1029/2001JB000172 (2002).
5. Fletcher, J. B. & Guatteri, M. *Geophys. Res. Lett.* **26**, 2295–2298 (1999).
6. Nadeau, R. M. & McEvilly, T. V. *Science* **285**, 718–721 (1999).
7. Harris, R. A. & Archuleta, R. J. *Geophys. Res. Lett.* **15**, 1215–1218 (1988).
8. Thurber, C. et al. *EOS Trans. Am. Geophys. Union Suppl.* **81**, S318 (2000); www.earthscope.org/safod.html

Neurobiology

Plasticity and the older owl

Hemai Parthasarathy

Age reduces the brain's ability to adapt to change. But a surprising measure of neural adaptation does occur, at least in one experimental situation, if change is introduced bit by bit.

They say that you can't teach an old dog new tricks. But on page 293 of this issue¹ Linkenhoker and Knudsen provide evidence that you can at least teach an old owl new tricks — if you train it in small steps. This result implies that the brains of older animals, including ourselves, may be capable of greater change in the form of adaptive plasticity than previously expected.

Owls are nocturnal hunters which rely on auditory as well as visual information to track their prey. Coordinated maps of auditory and visual space are formed in the optic tectum, the midbrain region responsible for orienting the owl to its target². The visual map is essentially a direct representation of information from the retina, but the auditory map results from a more complex computation, based in part on the microsecond differences between the time of arrival of sounds at each ear. For proper coordination of its auditory and visual worlds, the young owl must learn, for instance, that when it is looking directly at a mouse, the mouse's terrified squeak reaches both ears at the same time. But when the squeak reaches the right ear first by a certain number of microseconds, the owl should expect to see the mouse a certain number of degrees to the right.

Experiments with juvenile birds have shown that visual information has the upper hand in maintaining the coordination between these two maps. In such birds, Knudsen and a colleague³ previously found that an optical shift of the visual world, achieved by placing prism glasses on the owls, results in a gradual shift of the auditory

map to keep sight and sound coordinated. This plasticity may seem somewhat maladaptive under the conditions of the experiment — that is, although it is the visual world that has changed the visual map stays the same. However, it implies that outside the laboratory, over the course of evolution, vision has proven to be the more accurate and reliable indicator of spatial location for guiding orienting behaviours.

Although adaptive adjustments in the brain's representation of the auditory world

occur with relative ease in juvenile birds, it was thought that adult owls were less malleable in this respect (Fig. 1). The same prisms placed on an older owl, even for a period of months, cause only a tiny shift of the auditory map. In line with the terminology used in other kinds of neural systems, this window for large-scale auditory–visual realignment might be called a 'critical period' in development.

Evidence for critical periods in neural development dates back to the classic experiments of Hubel and Wiesel⁴, who deprived kittens of visual input to one eye and then showed that the animals' binocular vision, as well as the responses of their visual-cortex cells, differed from those of kittens raised normally. But identical manipulation of visual information had no effect in older cats. Other familiar examples of critical periods include language acquisition in humans and song learning by certain birds⁵. Critical periods in cognitive development are the rationale for playing Mozart to babies.

Linkenhoker and Knudsen¹ now show that the auditory map in older owls does indeed cope poorly with a large shift of the visual world. However, by shifting the visual world gradually, in several small steps, the map achieves a much greater degree of adaptive change than was previously believed possible. If, instead of being exposed to a single prismatic shift of 23°, older owls were first exposed sequentially to shifts of 6°, 11° and 17°, they acquired new auditory maps that could be reproduced by subsequent exposure to a single large prismatic shift.

These results by no means overturn the principle of a critical period. The auditory maps in juvenile owls adjust to much larger changes, and to about twice the extent, compared with adult owls. However, these

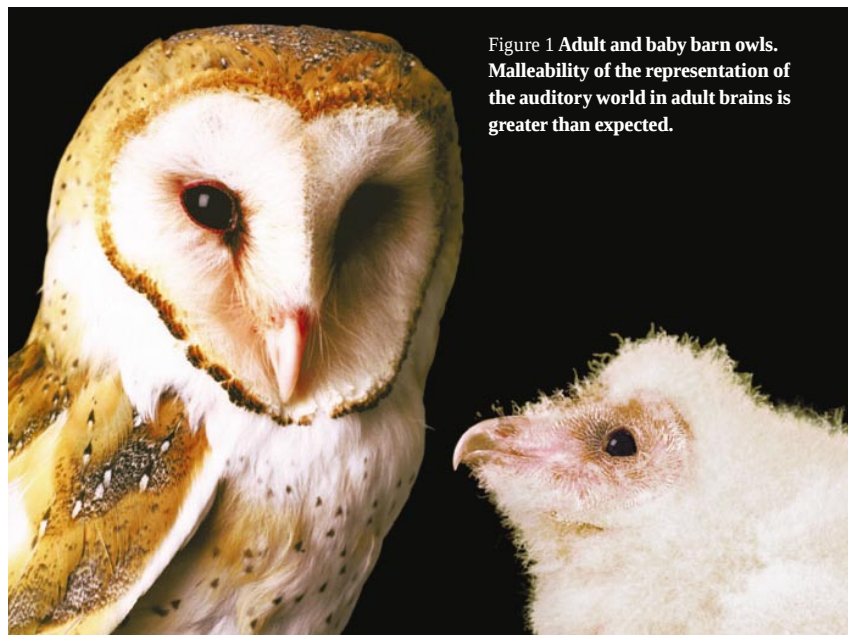


Figure 1 Adult and baby barn owls. Malleability of the representation of the auditory world in adult brains is greater than expected.

ANNE KNUDSEN & MELANIE CIFERRI

findings do challenge us to be more precise in defining what is meant by a critical period. They also challenge the notion that the same training regimes should be used when assessing adult and juvenile capacities for neural plasticity. Although there are striking parallels between the mechanisms of plastic change that operate during development and in adulthood, it is by no means clear that they will be the same, even within one system. A better understanding of the neural mechanisms underlying this form of plasticity in adult owls should lead to a better grasp of the limits of, and capacities for, adult learning.

There is some evidence that the degree of plasticity in this system may be limited by the extent of branching of the relevant neuronal elements, which seems to shrink as the owl matures. Knudsen's group⁶ has shown that large prismatic shifts can induce adaptive change in the auditory maps of adult owls that had previously adapted to the same large shifts as juveniles. This suggests that learning during the juvenile period causes some neuroanatomical or neurochemical foundation to be laid down in these animals that enables the signal for change to manifest itself over a much wider range than would otherwise be seen in the adult. It is tempting to think that the training regime used in the current experiments was more successful because it

was working within the constraints of more refined neural architecture that could bootstrap a more gradual, progressive change.

Finally, it should be noted that these experiments were done in laboratory owls that weren't really doing what owls like to do best with their orienting behaviours — hunt for prey. The plasticity measured here is in terms of days and weeks of experience, rather than, for instance, the number of practice trials, which is often considered the important variable in perceptual learning experiments in humans. An owl whose survival depends on successfully orienting to its prey may display an even greater capacity for change, for instance by virtue of neuromodulatory systems that become engaged during such 'arousing' activities as hunting. Whatever the case, these experiments should offer the hope to those of us who aren't getting any younger that improving learning ability just means finding new strategies for inducing plasticity in our brains. ■

Hemai Parthasarathy is a senior editor at Nature.

1. Linkenhoker, B. A. & Knudsen, E. I. *Nature* **419**, 293–296 (2002).
2. Knudsen, E. I. *Nature* **417**, 322–328 (2002).
3. Knudsen, E. I. & Knudsen, P. F. *J. Neurosci.* **9**, 3297–3305 (1989).
4. Hubel, D. H. & Wiesel, T. N. *J. Physiol. (Lond.)* **206**, 419–436 (1970).
5. Doupe, A. J. & Kuhl, P. K. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
6. Knudsen, E. I. *Science* **279**, 1531–1533 (1998).

Gamma-ray bursts

Light from darkness

Gerald J. Fishman

Telescopes must respond quickly to pinpoint transient gamma-ray bursts and pick up their afterglow. Mysterious 'dark' bursts seemed to produce no light at optical wavelengths, but a faint signal has now been detected.

Gamma-ray bursts (GRBs) are intense flashes of radiation that originate in the furthest reaches of the Universe. The bursts are expected to be followed by an afterglow of radiation at optical wavelengths. But for many GRBs (around 60% of the total sample of less than 40 bursts that have now been followed up), this afterglow seems to be missing. Hence, they have become known as 'dark' bursts. In a paper to appear in the *Astrophysical Journal*, Berger *et al.*¹ report data from an armada of telescopes that reveal a faint optical afterglow from a GRB that would otherwise have been classed as 'dark'. Their observations suggest that the seeming lack of optical emission from dark bursts is in fact a result of insufficiently prompt and sensitive observations.

Astronomers had thought that GRBs occurred within our Galaxy, but in the early 1990s it was realized that these explosive events actually occur at cosmological (extragalactic) distances, of the order of 15 billion light years away. The distance issue was settled by an observational breakthrough — the

accurate and rapid location of GRBs by the Italian–Dutch spacecraft Beppo–SAX. Following the initial detection and accurate location of gamma-rays by Beppo–SAX, the optical-radiation counterpart could be determined by other telescopes. This also makes it possible to measure the distance of the burst from the Earth. More importantly, detailed follow-up observations could then be made of the burst's afterglow and its host galaxy at other wavelengths by some of the most powerful telescopes in the world, both ground-based and space-based (in particular, the Hubble Space Telescope and the Chandra X-ray Observatory).

The field of prompt GRB follow-up observations has burgeoned in the past four years, in both the observational and the theoretical areas. Afterglow observations are being made at X-ray, optical, microwave, infrared and radio wavelengths. In each observational band, and by combining data from different bands, astronomers can make clever use of the data, such as providing a measure of the angular size of the emitting

region by observing radio variability².

For gamma-ray bursts, the observed optical (and X-ray) emission falls off rapidly with time. Perhaps in no other area of astronomy are observations needed so rapidly — ideally, minutes to a few hours following the burst. Any given telescope on the Earth has only about a 10% chance of making a quick observation, considering its location and the variety of limitations with which it must contend. The result described by Berger *et al.*¹ was made possible by a rapid response from the HETE spacecraft, in concert with an interplanetary network of spacecraft³ in orbit around Earth and Mars, and beyond. But such opportunities are rare.

The optical emission from the burst studied by Berger *et al.* (GRB 020124) decayed even more rapidly than most gamma-ray bursts, and it resided in the faintest host galaxy seen so far. It seems that the characteristic faintness and rapid decay of dark bursts is an intrinsic property of the burst, rather than the result of some other possible, external condition such as dust in the host galaxy. Berger *et al.* conclude that GRB 020124 is intrinsically faint at optical wavelengths and that it would have been classified as a dark GRB were it not for the prompt observations. Although the authors are cautious about drawing strong conclusions from this single burst, their result provides a new constraint on the models of bursts and afterglows that may render some of them untenable.

Theorists are thrilled to be guided and constrained by the new observational data. Most of them agree that a GRB is generated by very high-energy electrons and positrons produced in an initial explosion, and collimated into a jet perhaps only a few degrees wide. But the origin of the explosion and the type of star or stellar system involved are still shrouded in mystery. In fact, there are indications that more than one type of explosion is at work in the GRB phenomenon.

The future looks promising for obtaining accurate positions for more GRBs, with accompanying follow-up observations of their afterglow. In August, HETE recorded another triumph, by pinpointing a GRB less than four hours after the burst, leading to a flurry of observations from major telescopes, and from amateur astronomers as well. This provided a wealth of data on the GRB afterglow, including radio observations, an optical polarization measurement and fast Hubble and Chandra follow-up observations. The distance of the GRB from the Earth was also determined within a day of its detection by HETE.

On a more negative note, however, Beppo–SAX ceased operation earlier this year and the HETE spacecraft, although still operating, has had limited success, producing only a few GRB locations in the past couple of years because of operational and programmatic problems. The interplanetary network of

spacecraft, exploited by Berger *et al.*¹ in conjunction with HETE, can itself provide accurate locations of GRBs, although usually only many hours after the event, because communication with the distant spacecraft is limited.

But in late 2003, NASA will launch the Swift spacecraft⁴, which is expected to locate around 150 GRBs per year. In 2006, the GLAST spacecraft⁵ will likewise detect numerous GRBs and will cover an unprecedented broad energy range in the X-ray and gamma-ray regions. Beyond that, a large, next-generation GRB mission⁶ is under study, which would cover almost the entire sky with a sensitivity perhaps ten times that of Swift. In addition, about a dozen robotic and automated telescopes are operational or being developed to catch the early optical emission

from GRBs. And last but not least, as has been proved on at least four occasions so far, a well-prepared and knowledgeable amateur astronomer with a modest-sized telescope can make a substantial scientific contribution to the field of GRB research. ■

Gerald J. Fishman is in the Space Science Department, SD50, NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA.

e-mail: jerry.fishman@msfc.nasa.gov

1. Berger, E. *et al.* *Astrophys. J.* (in the press); Preprint astro-ph/0207320, <http://arXiv.org>
2. Waxman, E., Kulkarni, S. & Frail, D. *Astrophys. J.* **497**, 288–293 (1998).
3. Hurley, K. *et al.* *Astrophys. J. Suppl.* **122**, 497–501 (1999).
4. www.swift.psu.edu
5. www-glast.sonoma.edu
6. Grindlay, J. *et al.* in *Gamma-Ray Burst and Afterglow Astronomy 2001: A Workshop Celebrating the First Year of the HETE Mission* (ed. Ricker, G.) (Am. Inst. Phys., in the press).

Developmental biology

Sharp peaks from shallow sources

Hans Meinhardt and Siegfried Roth

The use of mathematical modelling to formulate and test theories is still quite rare in biology. It has now been applied to show how a robust, sharp peak of signalling molecules can be formed in developing fruitfly embryos.

It was proposed many years ago that, during animal development, graded concentration profiles of signalling molecules could provide information telling cells where they are in the embryo and what they should do. Since then, many examples of such molecules — known generically as morphogens — have been found^{1,2}. Concentration gradients can be generated through the production of morphogens at a particular source, followed by their diffusion and degradation in appropriate regions. Often, sharp peaks occur in the distribution

of morphogens, which might suggest that they require a sharply localized source. But in fact nature has circumvented this requirement, as can be seen from the development of the dorsal–ventral (back-to-belly) body axis in insects and vertebrates, where the source is plateau-like. The question of how this leads to a sharp concentration peak is difficult to tackle by traditional experimental means, as the system contains many variables. On page 304 of this issue, Eldar and colleagues³ describe how they used mathematical modelling to address the problem, and carried

out experiments to back up their predictions.

The authors³ were looking at the development of the dorsal–ventral pattern of fruitfly embryos. The embryo first becomes polarized by a gradient of the Dorsal protein, with highest levels at the ventral side, defining the ventral midline of the embryo. (The name of the protein, Dorsal, refers to what happens if it is mutated — the embryo has no ventral pole and forms only dorsal structures.) The subdivision of the embryo into dorsal and ventral regions then occurs through the activation or repression of target genes, in a manner that depends on the concentration of Dorsal.

Some of these target genes are needed to create another gradient — in this case, of activated morphogens from the bone morphogenetic protein (BMP) family. Highest levels of active BMPs will be at the dorsal side, defining the embryo's dorsal midline. One of the target genes encodes the BMP called Decapentaplegic (Dpp), which is expressed in a large dorsal domain. A second BMP — Screw (Scw), which cooperates with Dpp — is expressed uniformly. A further target of Dorsal is the BMP inhibitor called Short gastrulation (Sog), which is expressed in ventral–lateral regions.

A first look at this system suggests that some sort of conventional source–sink mechanism is at work: overall, the BMPs are mainly produced in the dorsal region (the source), and are antagonized by Sog spreading from the ventral site (the sink), leading to graded BMP activity⁴. But matters aren't quite so simple. First, Sog has not only a short-range inhibitory effect on BMP activity, but also a long-range enhancing effect^{5,6}. And second, the BMPs and Sog are all produced in broad rather than sharply focused regions, so one would expect the BMP gradient also to have a plateau-like distribution. Yet in fact the levels of active BMPs peak

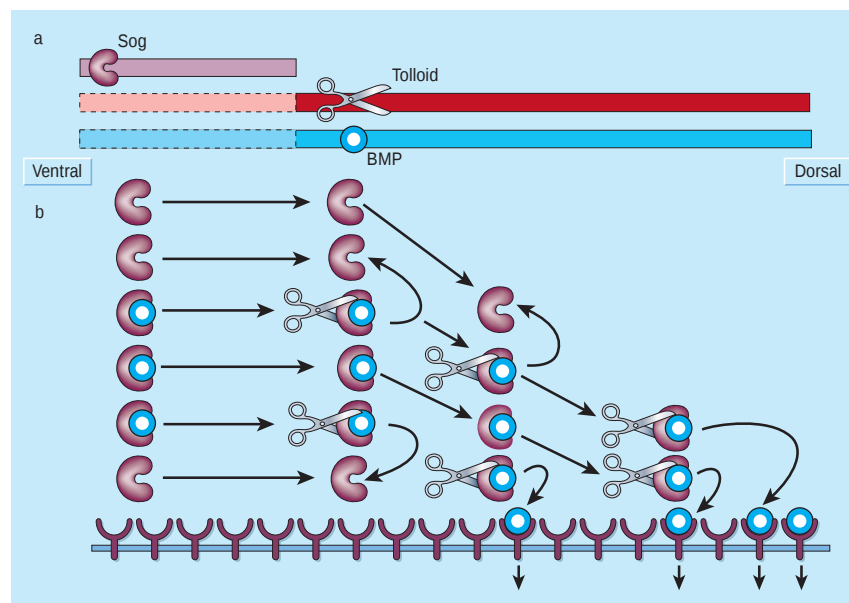


Figure 1 Model for generating a sharp gradient of bone morphogenetic proteins (BMPs) in fruitfly embryos, according to Eldar *et al.*³. a, Bars indicate the expression domains of BMPs, their inhibitor Sog, and the Sog-cleaving enzyme Tolloid. In this model, Sog needs to be produced ventrally. Tolloid is usually expressed dorsally, but Eldar *et al.* find that forced uniform overexpression (dashed bar) does not interfere with gradient formation. The BMP Scw is produced evenly, and the BMP Dpp is produced in the same domains as Tolloid. Again, uniform Dpp production does not affect gradient formation¹⁴. b, Eldar *et al.* find that two conditions must be met to ensure gradient formation. First, BMPs can move only when in complex with Sog. This leads to ventral-to-dorsal transport of BMP–Sog complexes. Second, Tolloid cleaves Sog only when Sog binds BMPs. Cleavage leads to release of BMPs, which bind either a new Sog molecule or the surface of nearby cells, leading to signalling (downward arrows). That will occur preferentially in dorsal positions, where Sog levels are low.

sharply in a stripe along the dorsal midline. Eldar *et al.*³ now offer an explanation for this phenomenon: they propose that Sog coordinates BMP transport with BMP release, and show by modelling that this produces a robust, sharp peak of active BMPs.

The authors began with computer simulations, based on a set of equations, in which they varied both the components and the parameters (protein concentrations and rate constants). To test their findings, they carried out experiments using Dpp. In essence, Eldar *et al.* show that BMPs by themselves probably have a very low diffusion rate. Only when bound to Sog do they become mobile, enabling them to be moved around the embryo. The enzyme Tolloid cleaves Sog, and Eldar and colleagues' models suggest that this happens preferentially when Sog is bound to BMPs, as proposed previously⁷. This releases the BMP molecules, which can act on cells in areas where the concentration of inhibitory Sog is low. That mainly happens at some distance from the

ventral source of Sog. In this way, a sharp focus of active BMPs can be built up at a distance from the source of Sog and on a plateau-like source of BMPs (Fig. 1).

Significantly, the authors show that this molecular network is robust to changes in the dosage of the genes involved. They first modelled a simplified system consisting only of Sog, Tolloid and one BMP. In this case, with certain parameters the system was insensitive to the gene dosage of all its components. If both Dpp and Scw were included, the system became unstable. Moderate stability was achieved if molecularly distinct complexes transported the two BMPs, with Sog alone transporting Scw, and both Sog and Twisted gastrulation (Tsg) transporting Dpp. This fits with experimental data^{6,8,9}. However, even in the presence of Tsg the system remained sensitive to the Dpp gene dosage, again in agreement with experiments. Whether dealing with the simple or the more complex system, the most remarkable outcome of the simulations is that

robustness to variations in system components is always linked to production of a sharp dorsal peak of BMP activity.

A similar mechanism is probably at work in vertebrates, where the same molecular players are involved¹⁰ — although in opposite regions¹¹. For instance, the Dpp homologues BMP2 and BMP4 are morphogens; they are produced in a broad ventral region, and are antagonized by the Sog homologue Chordin, localized dorsally in the so-called Spemann organizer. Homologues of Tolloid and Tsg^{8,9} are also found in vertebrates.

The puzzle that remains is why such a sophisticated mechanism is needed. At large distances from a localized source or sink, a gradient may be shallow and thus inappropriate for reliable subdivision. In many cases this problem is circumvented by using two gradients with opposite slopes, generated from independent sources at each end of the axis. An example is the fruitfly head-to-tail (anterior–posterior) axis, which is set up by anterior and posterior morphogen gradients that form largely independently of each other¹². Why isn't dorsal–ventral patterning achieved in a similar way?

The answer may lie in a peculiarity of the dorsal–ventral pattern: the initial peak of Dorsal protein must have the geometry of a long, narrow stripe rather than a small patch. The generation of such a pattern requires a complex set of interactions; otherwise, the stripe would be crooked, or decay into separated patches, or bifurcate. Correct patterning is achieved in vertebrates and insects in different ways — in vertebrates by a sequential elongation of the midline by the organizer, in insects by a repulsive effect from the dorsal side, which orients the Dorsal gradient and localizes the ventral midline¹³. Maybe the effort involved in generating one such midline was so great that the Sog–Tsg–BMP system was developed to use the information contained in that midline to create another at the opposite pole. ■

Hans Meinhardt is in the Department of Evolutionary Biology, Max-Planck-Institut für Entwicklungsbiologie, Spemannstrasse 37–39/IV, D-72076 Tübingen, Germany.

e-mail: hans.meinhardt@tuebingen.mpg.de

Siegfried Roth is at the Institute for Developmental Biology, Universität zu Köln, Gyrhofstrasse 17, 50923 Köln, Germany.

e-mail: siegfried.roth@uni-koeln.de

1. Gurdon, J. B. & Bourillot, P.-Y. *Nature* **413**, 797–803 (2001).
2. Lander, A. D. *et al.* *Dev. Cell* **2**, 785–796 (2002).
3. Eldar, A. *et al.* **419**, 304–308 (2002).
4. Srinivasan, S. *et al.* *Dev. Cell* **2**, 91–101 (2002).
5. Ashe, H. L. & Levine, M. *Nature* **398**, 427–431 (1999).
6. Decotto, E. & Ferguson, E. L. *Development* **128**, 3831–3841 (2001).
7. Marques, G. *et al.* *Cell* **91**, 417–426 (1997).
8. Ross, J. J. *et al.* *Nature* **410**, 479–483 (2001).
9. Larraín, J. *et al.* *Development* **128**, 4439–4447 (2001).
10. Dale, L. & Wardle, F. C. *Semin. Cell Dev. Biol.* **10**, 319–326 (1999).
11. Arendt, D. & Nübler-Jung, K. *Nature* **371**, 26 (1994).
12. St Johnston, D. & Nüsslein-Volhard, C. *Cell* **68**, 201–219 (1992).
13. Meinhardt, H. *BioEssays* **24**, 185–191 (2002).
14. Jazwinska, A. *et al.* *Development* **126**, 3323–3334 (1999).

Cell biology

Proteins tracked in a flash

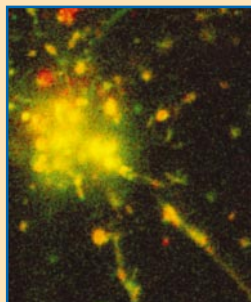
Cell biologists have a lot to thank jellyfish for. The species concerned has provided them with green fluorescent protein (GFP), a tool for monitoring the dynamics of gene expression and following the movement of proteins and even entire cells. Over the years, the usefulness of this 'molecular reporter' has been boosted by the isolation of GFP variants possessing greater fluorescence, altered excitation and emission wavelengths, and greater stability in living cells.

But there was scope for further improvement. In particular, tracking protein movement in real time was difficult if the tagged protein became evenly distributed in the cell under steady-state conditions. Writing in *Science* (**297**, 1873–1877; 2002), George H. Patterson and Jennifer Lippincott-Schwartz now describe a possible solution to this problem. They have produced a GFP variant that shows greatly increased fluorescence if it is activated by light; and it functions in physiological conditions. Using this photoactivatable GFP, tagged proteins in a small area of the cell can be selectively marked by light activation, so

that their movement through the cell can be followed against a dark background by fluorescence microscopy.

Normal GFP contains a mixed population of neutral and anionic chromophores — molecules that absorb light then re-emit it. These are respectively associated with a major light-absorption peak at a wavelength of 397 nm and a minor peak at 475 nm. Intense illumination at 400 nm shifts the population to the anionic form, thereby increasing the absorbance of the minor peak. Patterson and Lippincott-Schwartz set out to develop a GFP variant that had a negligible 475-nm peak. They reasoned that illumination at 400 nm would then produce a much greater proportional increase in 475-nm absorbance compared with the normal protein, and therefore increase optical contrast.

One variant, called PA (for 'photoactivatable') GFP, produced a 100-fold increase in fluorescence at 488 nm *in vitro*, and was stable under various conditions. Patterson and Lippincott-Schwartz then tested this variant in living cells. They found that selective



photoactivation of PA-GFP in either the nucleus or cytoplasm of the cell led to a brightly fluorescent population that moved rapidly between the two areas. The authors also attached PA-GFP to a marker protein called lgp 120, found in a cellular compartment, the lysosome, that is responsible for digesting unwanted material. Before photoactivation, little fluorescence of lgp 120 was seen. But after photoactivation, lgp 120 was evident in most of the lysosomes within 20 minutes (see figure), showing that rapid exchange occurs between these organelles.

This new type of GFP should prove valuable for studying the temporal and spatial dynamics of proteins in the living cell. The future for such research literally looks brighter.

Deepa Nath

Physiology

Unhealthy surprises

Dante R. Chialvo

Fluctuations in heart rate can signal disease and may prove fatal. A measurement, based on entropy, of how 'surprising' the beat irregularity is, distinguishes healthy hearts from those suffering common forms of illness.

Although unnoticed, the healthy heart beats irregularly. The intervals between beats fluctuate widely, following complicated patterns reminiscent of music¹. The origin, significance and diagnostic value of this variability of heart rate is a subject of sustained interest for biologists and, more recently, for physicists also. Writing in *Physical Review Letters*, Madalena Costa and colleagues² analyse the statistics of a sample of heart beats. They find that healthy people and patients with heart conditions can be consistently differentiated by a surprisingly simple measure based on a thermodynamical concept: entropy.

The problem Costa *et al.* tackle is that of how best to describe a discrete series of consecutive events, in this case the time series of consecutive beat-to-beat intervals. There are already many ways of addressing this problem, but no single one seems able to solve it. Like others before them, the authors have used entropy to capture the complexity of the time series. Entropy estimates the average uncertainty of a series of discrete events. In the case of the heart, it estimates an average of how much the next beat-to-beat interval will surprise us. If the intervals are equally spaced, and thus unsurprising (for example, the unrealistic case of a perfect clock), entropy is a minimum. At the other extreme, if the beat-to-beat intervals are randomly distributed over a wide range, each event will not be easily anticipated and entropy increases.

A straightforward measurement of entropy can capture some of the predictability of heart-rate variability, but has limitations. For instance, in conditions such as atrial fibrillation (twitching of the upper chambers of the heart) that produce highly erratic rate fluctuations, entropy is actually comparable to that in perfectly healthy hearts. Therefore, researchers are adding new quantities derived from entropy in their attempts to find better ways of teasing apart different physiological and pathological states.

Costa *et al.*² were inspired by the fact that physiological time series often exhibit novel features at different timescales. They hypothesized that entropy could vary according to the timescale at which it is measured. By re-sampling the original time series at various scales — from considering the sample as a whole, to dividing it into up to twenty sub-samples — they obtained a

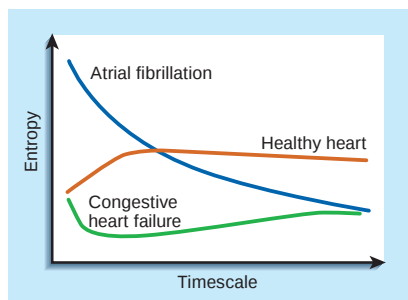


Figure 1 Entropy and the healthy heart. By analysing the entropy, or disorder, of the intervals between heart beats on different timescales, Costa *et al.*² show that the entropy curve distinguishes the behaviour of a healthy heart from that of an unhealthy heart, for example one suffering from atrial fibrillation or congestive heart failure.

collection of data possessing different coarse-graining from which they calculated a measure of entropy. The resulting quantity, termed 'multiscale entropy', showed that the heart-beat time series of healthy people asymptotically approaches a constant value of entropy as the measurement scale is increased (Fig. 1). Intuitively, this is what one would expect, because, as in the case of music, in healthy beat-to-beat time series new features come into play at each timescale. Consequently, regardless of the length of the intervals at which the series is observed, the uncertainty remains constant. On the other hand, the analysis of heart patients shows a departure from this 'constancy of surprise', with distinctive multiscale-entropy patterns. In patients with atrial fibrillation, entropy decreases, but the opposite is seen in patients with cardiac failure (Fig. 1). Thus, multiscale entropy finds differences that are missed by other approaches.

Although most efforts, including those of Costa *et al.*, are concerned with the statistics of thousands of beats, work is also being done on the dynamics of just a few beats. The precise sequence in which the heart fluctuates is very informative. For instance, it is known that the length of two consecutive beat-to-beat intervals cannot be made arbitrarily different, because it is limited by the intrinsic dynamics of the heart's own pacemaker. Think of a swing in motion, for example: if you try to speed it up or slow it down with isolated pushes, you realize that

two consecutive oscillations cannot be made arbitrarily very different. In fact, the degree to which an oscillator can be advanced or delayed by a single perturbation is a dynamical signature of the system and has been used to predict many aspects of dynamics³.

This reasoning can be reversed: by analysing the dynamics of just a few consecutive beat-to-beat intervals, the status of the system can be inferred. Tatenno and Glass⁴ showed that this can reveal important information. They developed a technique that could detect atrial fibrillation in patients, based on an analysis of fifty consecutive beat-to-beat intervals. Histograms generated from the differences between consecutive intervals had characteristic shapes that were dependent on the patient's heart rhythm. By comparing a patient's histogram with standard histograms, Tatenno and Glass were able to identify with high sensitivity and specificity whether the patient's rhythm matched that of atrial fibrillation or some other form of arrhythmia. Atrial fibrillation is a common problem, associated with a high risk for stroke, so algorithms that can detect it, especially if they can be incorporated in portable devices, should be useful for optimizing the administration of drugs and other treatments.

Neither advance, whether achieved by studying a few heart beats or Costa and colleagues' investigation of many heart beats², touches on the physiological mechanisms that generate the fluctuations. It is well known — though often not well appreciated — that if one variable in an organism is to be kept relatively constant, something else must fluctuate. To maintain a relatively constant supply of nutrients, the heart rate (as well as other cardiovascular variables) must somehow adjust to variations in the demand. This might explain most heart-rate fluctuations. Of course, it is not known precisely how this happens, nor what mechanisms are responsible for the quantitative features of the fluctuations. A challenging task will be to create a model that, without writing the fluctuations in explicitly, would still be able to reproduce beats possessing all the music we see in healthy hearts. Until the correct statistical behaviour arises out of the dynamics of a physiological model, our understanding of this riddle will not be complete. ■

Dante R. Chialvo is in the Department of Physiology, University of California at Los Angeles, California 90095, and the Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021, USA.
e-mail: dchialvo@ucla.edu

1. Appel, M. L., Berger, R. D., Saul, J. P., Smith, J. M. & Cohen R. J. *J. Am. Coll. Cardiol.* **4**, 1139–1148 (1989).
2. Costa, M., Goldberger, A. & Peng, C.-K. *Phys. Rev. Lett.* **89**, 068102 (2002).
3. Glass, L. & Mackey, M. *From Clock to Chaos: The Rhythms of Life* (Princeton Univ. Press, New York, 1988).
4. Tatenno, K. & Glass, L. *Med. Biol. Eng. Comput.* **39**, 664–671 (2001).

Conservation biology

Science, sex and the kakapo

William J. Sutherland

Sex-allocation theory predicts how the sex ratio of offspring should vary with the mother's physical condition. Applying this theory has helped in retrieving a charismatic parrot from the edge of extinction.

The kakapo, a nocturnal, flightless parrot found in New Zealand, is an endangered species, its rarity and eccentricity making it a high-profile test for conservationists. At a meeting last month*, Mick Clout (Univ. Auckland) reported a scientific success story in reducing the bird's risk of extinction. Providing extra food to improve fledging success seems to have inadvertently resulted in breeding females producing an excess of male young, as predicted by evolutionary theory^{1,2}. Given this finding, conservation workers decided to withhold extra food until after the females had laid their eggs, resulting in a much-needed boost in the number of female fledglings this year.

The kakapo (*Strigops habroptilus*; Fig. 1) was once common in the three largest islands of New Zealand, but its distribution shrank following the spread of black rats and stoats introduced from Europe in the 1870s and 1880s. By the 1950s, kakapo were extinct on North Island and only 18 remained in a remote and mountainous part of South Island. The stoats are thought to have eaten not only eggs and chicks, but also incubating females, so all 18 survivors were male. The only remaining wild females were on Stewart Island (south of South Island), which stoats had not colonized, although rats and feral cats were present. A bold decision, taken in 1982, was to capture all of the remaining birds and release them on predator-free

islands, including Codfish Island, off the coast of Stewart Island. A world population of 62 birds in 2001, with only 21 adult females, created an urgent need for more females, yet only six female fledglings had been produced since 1982.

In most years, kakapo feed mainly on foliage and other poor-quality foods that are insufficient to allow breeding. Egg-laying is restricted to those years in which certain trees — podocarps — undergo mass fruiting, known as masting. Since 1989, supplementary food, such as nuts and sweet potato, has been given to some individual kakapo from dispensers in their home ranges, resulting in a 15% increase in body weight for fed birds. Unlike natural masting, supplementary food did not seem to induce breeding. But fed females spent less time off the nest, which increased chick survival.

The theory of sex allocation proposed by Trivers and Willard¹ states that females in better physical condition should produce more offspring of the sex that shows the greater benefit from the improved condition. Kakapo reproduce in leks — communal display grounds in which males fight to defend small territories within which they display vigorously. Females visit just to select a mate before raising the young alone elsewhere. Studies of various lek-breeding species show strong competition between males, with a few high-quality individuals obtaining almost all the matings³.

Male kakapo fight, sometimes to the



100 YEARS AGO

I have frequently observed with interest the erect attitude assumed by the small Agamid lizard *Otocryptis bivittata*, Wieg., when running rapidly, and have long suspected that the short front legs were not used at such times. But the rapidity with which the animal runs, and the nature of the ground which it usually frequents, have prevented very close observation. I have, however, recently fully satisfied myself that its action is truly bipedal... On several occasions one of them has crossed a smooth sanded road immediately in front of me. I have thus been able to see clearly that the anterior limbs are carried quite free from the ground, progress being effected solely by the long hind limbs.

ALSO...

According to a Reuter telegram from Rome, the Italian postal authorities have examined a scheme submitted by an engineer, named Piscicelli, for the establishment of an electric postal service. It is proposed, by means of this system, to transmit letters in aluminium boxes, travelling along overhead wires at the rate of 400 kilometres an hour. A letter could thus be sent from Rome to Naples in twenty-five minutes and from Rome to Paris in five hours.

From *Nature* 18 September 1902.

50 YEARS AGO

The dusky-footed wood rat (*Neotoma fuscipes* Baird.) is a medium-sized rodent native to North America... Wood rats are nocturnal and are seldom seen without special search, but the conspicuous houses in which they live, built of sticks above ground, readily show their presence... A wood rat's ability to survive is affected by the conditions that enable it to leave its home and those that affect it away from shelter. A rat that lives in a house providing protection from the weather and predators is exposed to these dangers when away from home. It leaves its house mainly to get food and housebuilding material, to escape from dangerous intruders or intolerant members of its own family, and to find other rats. Many rats contribute to the maintenance of few houses, which last much longer than any single occupant, but the character of each house is dictated by its situation and the available materials; it remains constant in type throughout successive tenancies.

From *Nature* 20 September 1952.

*International Ornithological Congress, Beijing, China, 11–17 August 2002.



Figure 1 The eccentric kakapo. This remarkable nocturnal species differs from all other parrots in many ways, including being flightless, mating in display grounds ('leks') and breeding only in the infrequent years when specific trees undergo mass seeding. The sex ratio of the offspring depends on the condition of the mother, which has considerable implications for conserving the species.

death⁴, and larger males are more likely to mate. Male kakapo also grow faster and larger than females, and are thus presumably more costly to raise⁵. So if Trivers and Willard's hypothesis applies, mothers in good condition should produce more sons. Females provided with supplementary food indeed produced significantly more sons (an average of 67% against 29% for birds not given supplementary food)². There is accumulating evidence that maternal condition regularly affects the offspring sex ratio in a range of species⁶, including tree swallows⁷, lesser black-backed gulls⁸, great reed warblers⁹ and another species of parrot¹⁰.

A profuse fruiting of podocarp trees was expected to trigger kakapo breeding on Codfish Island this year, so the New Zealand Department of Conservation concentrated all remaining 21 adult females there. Supplementary feeding was delayed until after egg production. The idea was to avoid boosting female condition too much before laying, to avoid a male bias in offspring. Extra food was still provided after laying to maximize fledging success.

The strategy seems to have worked: as Clout described at the meeting, 15 of the 24 young fledged this year were female. This is a neat example of how behavioural ecology can benefit conservation, although the

mechanism by which females manipulate the sex ratio remains a mystery.

A common problem in conservation is that endangered species are pushed into sub-optimal areas as a result of predation by introduced species or exploitation in their core range. Among the bird species that have only really flourished once they were moved are the nene in Hawaii, the Lord Howe woodhen on Lord Howe Island, the takahe in New Zealand and the red kite in Britain. Intensive manipulation of the kakapo has had its successes, but the long-term solution will surely lie in finding ways of restoring predator-free areas on the main islands of New Zealand. ■

William J. Sutherland is at the Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

e-mail: w.sutherland@uea.ac.uk

1. Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).
2. Clout, M. N., Elliot, G. P. & Robertson, B. C. *Biol. Conserv.* **107**, 13–18 (2002).
3. Höglund, J. & Alatalo, R. V. *Leks* (Princeton Univ. Press, 1995).
4. Clout, M. N. & Merton, D. V. *Bird Conserv. Int.* **8**, 281–295 (1998).
5. Powesland, R. G. *et al.* *Ibis* **134**, 361–373 (1992).
6. Tella, J. L. *Trends Ecol. Evol.* **16**, 76–77 (2001).
7. Whittingham, L. A. & Dunn, P. O. *Mol. Ecol.* **9**, 1123–1129 (2000).
8. Nager, R. G. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 570–573 (1999).
9. Westerdaal, H. *et al.* *Behav. Ecol. Sociobiol.* **47**, 312–318 (2000).
10. Heinsohn, R., Legge, S. & Barry, S. *Proc. R. Soc. Lond B* **264**, 1325–1329 (1997).

Chemical physics

A delayed reaction

David E. Manolopoulos

The simple hydrogen exchange reaction was well understood until an unexpected effect emerged in detailed experimental measurements of the process. An explanation for this effect has now been found.

The scattering of a hydrogen atom from a hydrogen molecule ($\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$) is the simplest possible chemical reaction, as it involves only three nuclei and three electrons. This reaction has served as a benchmark system for understanding chemical reaction dynamics for nearly a century, and quantum-mechanical calculations of the electronic aspects of the process have been made with unrivalled accuracy. As a result, quantitative agreement between theory and experiment for the related reaction involving a deuterium molecule, $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$, was achieved in the middle of the last decade¹.

But more refined experiments and improved theoretical methods have since raised an intriguing question about the dynamics of this reaction at the quantum level^{2,3}. Although the reaction is dominated by a direct recoil mechanism — the incident hydrogen atom recoils along its original path after abstracting a deuterium atom to form the HD product molecule — some of the more subtle aspects of the reaction dynamics

revealed by recent experiments have been traced to a second, slower mechanism that occurs with a time delay of around 25 femto-

seconds². The question concerns the physical origin of this time delay, which has become the subject of some debate^{3,4}. On page 281 of this issue, Harich *et al.*⁵ provide an explanation.

In the $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$ reaction, effects beyond the direct recoil mechanism are expected to become apparent only when the quantum states of the reactant and product molecules are directly interrogated. A useful, measurable quantity is the 'state-to-state differential cross-section' — the probability for the reaction to proceed from reactants in a specific initial quantum state to products in a specific final quantum state as a function of the scattering angle between them. This quantity was first measured for the $\text{H} + \text{D}_2$ reaction at a single collision energy by Welge and co-workers¹ in the 1990s.

Following the suggestion of Miller and Zhang⁶, Althorpe *et al.*² have since measured the state-to-state differential cross-section over a range of collision energies, in the hope of gaining greater insight into the reaction dynamics of the $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$ process. By varying the wavelength of the photolysis laser used to photodissociate HBr precursor molecules (and hence generate the reagent H atoms), they were able to measure the differential cross-section for collision energies from 1.39 to 1.85 eV. Their reward was the observation, at a collision energy of around 1.64 eV, of an intriguing forward-scattering peak for the HD product molecule in a quantum state with a vibrational quantum number of three and a rotational angular momentum quantum number of zero. A theoretical analysis by the same authors² of this forward scattering showed that it was delayed relative to the backward-scattering recoil mechanism by some 25 femtoseconds, but their analysis did not produce a physical explanation.

In fact, there are two conceivable explanations for such a time delay³. The more obvious is a simple 'threshold' effect: as the reactants approach a potential-energy

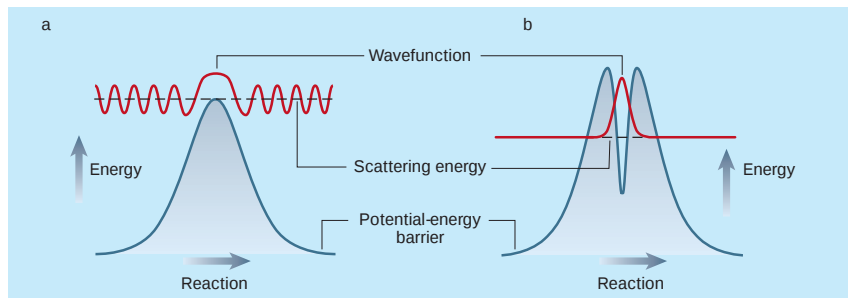


Figure 1 Wavefunctions and barriers. There are two possible explanations for the time delay seen in forward scattering in the reaction $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$. **a**, The scattering energy (dashed black line) coincides with the top of a potential-energy barrier (solid blue line); crossing this 'threshold' causes the reactants to slow down and leads to the increased amplitude of the quantum-mechanical wavefunction (solid red line) at the barrier maximum. **b**, A 'quasi-bound' quantum state exists in the well between the double maxima of a potential-energy barrier (solid blue line) before decaying into reaction products. The highly localized wavefunction (solid red line) indicates a 'resonance' effect. Harich and co-workers' analysis⁵ of wavefunctions in the $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction suggests that the time delay is due to the threshold effect, not the resonance process.

barrier they lose classical kinetic energy and slow down. This situation is depicted in Fig. 1a, which shows the symmetric scattering wavefunction for a one-dimensional potential-energy barrier when the reactant energy is equal to that of the barrier maximum. The slowing down of the reactants in classical mechanics is reflected by an increase in the amplitude of the quantum-mechanical wavefunction ψ , and hence an increase in the probability $|\psi|^2$ of finding the system in the region of the barrier maximum. But because the reactants simply slow down at the top of the barrier rather than becoming trapped there, the wavefunction associated with this reaction threshold effect still has significant amplitude away from the barrier maximum.

The second explanation does not have a classical analogue, and is therefore a little more exotic. It is possible for the reactants to become trapped on the potential-energy surface, forming a 'quasi-bound' quantum state. This situation also leads to a time delay, owing to the time it takes for the quasi-bound state to decay into reaction products, and it is known as a quantum reactive scattering resonance. The simplest example of such a resonance is illustrated in Fig. 1b, which shows the computed scattering wavefunction at the energy of a quasi-bound state supported by the well between the two potential-energy maxima of a double barrier.

Although these two explanations are not unrelated, because one situation can be deformed into the other by changing the potential-energy surface⁷, they are mathematically different^{8,9}, and the physical implications of this mathematical difference are too significant to regard the two situations as manifestations of the same phenomenon. Furthermore, it is clear from Fig. 1 that the scattering wavefunction behaves very differently in the threshold and resonance situations.

This last distinction has been exploited by Harich *et al.*⁵ to elucidate the origin of the time delay associated with forward scattering in the hydrogen exchange reaction. Rather than study the H + D₂ version of the reaction, they used a crossed molecular beam apparatus to study the related H + HD → H₂ + D reaction, at a collision energy of 1.2 eV. State-to-state differential cross-sections were measured by Rydberg tagging the product deuterium atom (and thereby inferring the quantum state of the hydrogen product molecule), a powerful and highly sensitive technique that was first used for the H + D₂ reaction by Welge and co-workers¹. Although the experiment was only done for a single collision energy, a distinct forward-scattering peak was seen in the differential cross-section of the H₂ product with a vibrational quantum number of zero and a rotational angular momentum quantum number of one.

This forward peak in the H + HD reaction has similar characteristics to that seen in the H + D₂ reaction by Althorpe *et al.*². Both

peaks occur for product states with low rotational quantum numbers, and the theoretical analysis of Harich *et al.*⁵ shows that the forward scattering in the H + HD reaction is again associated with a time delay, in this case of around 20 femtoseconds. But this analysis goes further and extracts the quantum-mechanical wavefunction that underlies the time-delayed reaction mechanism. The wavefunction turns out to be like the one shown in Fig. 1a, so Harich *et al.* conclude that the time delay in forward scattering in the H + HD reaction is caused by a reaction threshold effect rather than a reactive scattering resonance. The same is likely to be true in the case of the H + D₂ reaction.

This is, in fact, the best possible outcome for the field. A genuine quantum reactive scattering resonance has recently been identified through a theoretical analysis of integral and differential cross-section measurements

on the F + HD → HF + D reaction¹⁰. Hence, we now have concrete examples of both phenomena — reactive thresholds and reactive resonances — and the effects they have on experimental observations of the dynamics of chemical reactions. ■

David E. Manolopoulos is in the Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK. e-mail: mano@physchem.ox.ac.uk

1. Schnieder, L. *et al.* *Science* **269**, 207–210 (1995).
2. Althorpe, S. C. *et al.* *Nature* **416**, 67–70 (2002).
3. Fernandez-Alonso, F. & Zare, R. N. *Annu. Rev. Phys. Chem.* **53**, 67–99 (2002).
4. Aoi, F. *et al.* *J. Chem. Phys.* **117**, 2546–2556 (2002).
5. Harich, S. A. *et al.* *Nature* **419**, 281–284 (2002).
6. Miller, W. H. & Zhang, J. Z. *H. J. Phys. Chem.* **95**, 12–19 (1991).
7. Friedman, R. S. & Truhlar, D. G. *Chem. Phys. Lett.* **183**, 539–546 (1991).
8. Seideman, T. & Miller, W. H. *J. Chem. Phys.* **95**, 1768–1780 (1991).
9. Sadeghi, R. & Skodje, R. T. *Phys. Rev. A* **52**, 1996–2010 (1995).
10. Liu, K., Skodje, R. T. & Manolopoulos, D. E. *Phys. Chem. Commun.* **5**, 27–33 (2002).

Gene regulation

Reviving the message

Walter Keller and Georges Martin

Studies of developmental regulators in worms and cell-cycle regulators in yeast have revealed a new family of enzymes, which may affect the fate of specific messenger RNA molecules.

Messenger RNA molecules are crucial intermediates between genes and their encoded proteins. When a gene is activated, enzymes produce an mRNA copy of it; this mRNA in turn provides a template for the production of proteins, which carry out specific tasks in the body. mRNAs consist of strings of nucleotides, and in our cells most mRNA molecules have a long tract of 'A's (adenosine nucleotides) at one end — the so-called 3' end. Such poly(A) tails seem to be required for every step in an mRNA's life, including its export from its site of production in the cell nucleus, translation into protein, and stability¹. On page 312 of this issue, Wang and colleagues² describe an unusual enzyme, important in the development of the nematode worm *Caenorhabditis elegans*, that they propose lengthens the poly(A) tails on certain mRNAs. What's interesting is that this work apparently defines an entirely new class of such enzymes, and has implications for developmental and cell biology.

Messenger RNA precursors first become 'polyadenylated' in the nucleus, during or shortly after gene transcription, in a reaction involving two coupled steps: cleavage of the RNA to form a new 3' end, then poly(A) addition by a poly(A) polymerase enzyme. Polyadenylation can also, however, occur outside the nucleus, in the cell cytoplasm. During the early embryonic development of many animals and the maturation of germ cells (eggs and sperm), transcription is

largely switched off. Elongating the short poly(A) tail of dormant cytoplasmic mRNAs provides a rapid way to stabilize and activate them³, allowing proteins to be produced without transcription. (mRNAs need a long poly(A) tail to function as efficient templates in translation.)

In *C. elegans*, the *gld-2* and *gld-3* genes control various aspects of germline development, including the mitosis/meiosis decision — simply put, whether germline cells multiply to produce other such cells, or generate eggs and sperm⁴. But exactly how the proteins encoded by these genes regulate developmental decisions has been unclear. This is what Wang *et al.*² set out to investigate, and they have found that these proteins form a cytoplasmic poly(A) polymerase with a difference.

Wang *et al.* started by inspecting the predicted amino-acid sequence of the GLD-2 protein, and discovered an immediate clue to its biochemical function. The protein contains a domain that is similar to a region in certain nucleotidyltransferase enzymes (Fig. 1, overleaf). These enzymes form a protein superfamily to which all eukaryotic poly(A) polymerases belong⁵. The authors also found that GLD-2 is located in the cytoplasm of germline and early embryonic cells, and interacts specifically with GLD-3, which itself belongs to a family of RNA-binding proteins (J. Kimble *et al.*, personal communication).

Given this cytoplasmic location of GLD-2,

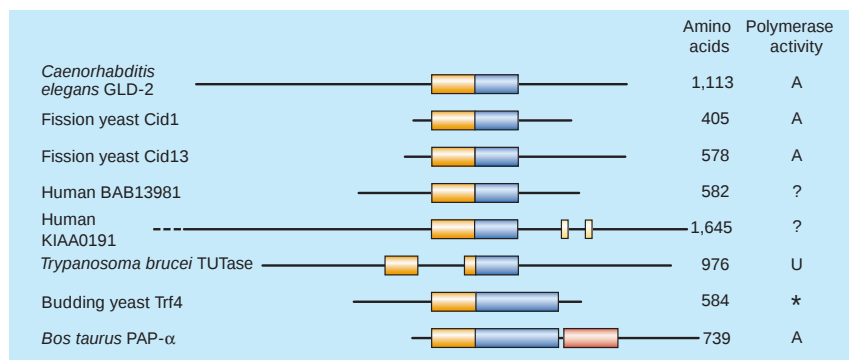


Figure 1 A new type of RNA-modifying enzyme? Wang *et al.*² have identified the GLD-2 protein as a poly(A) polymerase enzyme *in vitro*; they propose that it extends the tract of adenosine nucleotides at the 3' end of certain messenger RNAs *in vivo*. Unlike many such enzymes, GLD-2 can work only with another protein, GLD-3, which binds RNA. Other such unusual poly(A) polymerases may exist, as can be seen in this comparison of GLD-2 with other proteins. A nuclear poly(A) polymerase, PAP- α , which can bind RNA without a helper protein, is also shown. The proteins are represented as lines, with conserved regions shown as cylinders (orange, catalytic domain; blue, central domain; red, RNA-binding domain; yellow, protein-interaction domains called zinc fingers). The polymerase activities are A, adenylating, and U, uridylylating; the asterisk indicates that poly(A) polymerase and DNA polymerase activities have been reported for this protein.

and the slight similarity of its amino-acid sequence to those of nuclear poly(A) polymerases, Wang *et al.* decided to test the protein for RNA-dependent poly(A) polymerase activity *in vitro*. They found that GLD-2 alone had low poly(A) polymerase activity, but was stimulated by GLD-3, which by itself was completely inactive. Analysis of the reaction products showed that GLD-2 alone extended an RNA 'primer' by only a few adenosines, whereas GLD-2 and GLD-3 together made tails of up to 30 adenosines. Two GLD-2 mutant proteins, one designed to abolish its predicted catalytic centre and the other to disrupt its binding to GLD-3, were inactive when tested either alone or with GLD-3. All known poly(A) polymerases comprise a single protein. So Wang

et al. have found a new type of cytoplasmic poly(A) polymerase, in which GLD-2 provides the catalytic subunit and GLD-3 contributes the RNA-binding function.

These findings raise several questions. Which mRNAs are the physiological substrates of the newly discovered enzyme during early development? Can GLD-2 interact with other RNA-binding proteins to expand its substrate repertoire—an idea proposed by Wang *et al.*? Are GLD-2 and GLD-3 enough to carry out the reaction *in vivo*, or are other factors involved? And does this newly discovered process share any components with previously described cytoplasmic poly(A) polymerases involved in development, or with the nuclear 3'-end-processing apparatus?

Wang and colleagues' work also has ramifications that go beyond the control of early development. The polyadenylation of cytoplasmic mRNAs appears to be widespread in eukaryotes, and GLD-2 may represent a new family of bipartite cytoplasmic poly(A) polymerases. For example, the fission-yeast proteins Cid1 and Cid13 are cytoplasmically located relatives of GLD-2, and have poly(A) polymerase activity *in vitro*^{6,7}. Like GLD-2, Cid1 is involved in controlling the cell-division cycle. Cid13 has been proposed to increase the pools of nucleotides needed for DNA replication, by extending the poly(A) tail of the mRNA encoding Suc22 — part of an enzyme involved in nucleotide synthesis. Moreover, it has been reported⁶ that Trf4, a relative of the Cid proteins that occurs in budding yeast⁸, has *in vitro* poly(A) polymerase activity (although this is controversial, and previous *in vitro* tests identified Trf4 as a DNA polymerase⁹). Many more members of the GLD-2 family may exist, as inferred from sequence comparisons (Fig. 1).

Further insights into GLD-2 can be

gleaned from a look at its amino-acid sequence. Mammalian and yeast nuclear poly(A) polymerases have a three-domain structure consisting of a catalytic portion, a central linker and an RNA-binding domain. GLD-2, like most of its close relatives, lacks an RNA-binding domain. Nonetheless, Wang *et al.*'s comparison of the sequences of GLD-2 and its relatives with that of mammalian nuclear poly(A) polymerase suggests that the structure of the catalytic domain and part of the central domain is similar, despite considerable sequence divergence (Fig. 2). The most prominent conserved features are three aspartate amino acids — which chelate divalent metal ions — in the active site, and several amino acids in the catalytic and central domains that help to bind ATP (see Fig. 2 on page 313).

Moreover, analysis of mutant forms of mammalian poly(A) polymerases suggests that amino acids in a loop near the ATP-binding pocket are needed to bind the 3' end of mRNAs (our unpublished results). The sequence of this loop is also seen in GLD-2 and its relatives, and in 3'-terminal uridylyl-transferase — an enzyme in trypanosomes that catalyses the addition of uridine nucleotides to the 3' ends of intermediates of RNA editing⁹. So this sequence seems to be a hallmark of enzymes that elongate single-stranded RNA substrates.

Lengthening the poly(A) tails of selected mRNAs at specific times can both counteract normal mRNA turnover and stimulate translation. This is useful, because it bypasses the requirement for transcription and RNA processing, for example in times of metabolic stress or when the genome is damaged or inactive. Moreover, this mechanism — like others that act on cytoplasmic mRNAs — is much more rapid than controlling transcription, particularly at large genes. Cytoplasmic polyadenylation would gain great versatility if, as Wang *et al.* suggest², different mRNAs could be targeted by using different RNA-binding proteins to recruit a poly(A) polymerase. Whether that happens remains to be seen, but this and other questions will keep many of us busy and excited for a long time.

Walter Keller and Georges Martin are at the Biozentrum of the University of Basel, CH-4056 Basel, Switzerland.

e-mail: walter.keller@unibas.ch

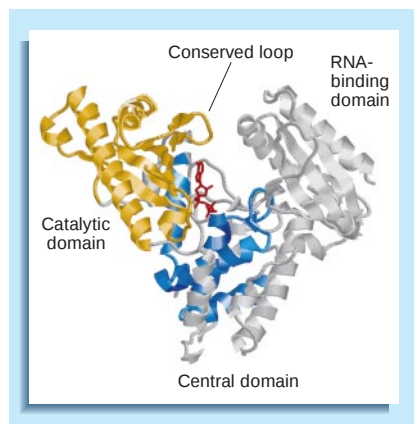


Figure 2 Structure of the mammalian nuclear poly(A) polymerase, with regions of similarity to GLD-2 and its relatives highlighted (colour code as in Fig. 1). The conserved loop might be a hallmark of enzymes that elongate single-stranded RNA substrates. ATP in the active site is shown in red. (Figure modified from ref. 10.)

- Shatkin, A. J. & Manley, J. L. *Nature Struct. Biol.* **7**, 838–842 (2000).
- Wang, L., Eckmann, C. R., Kadyk, L. C., Wickens, M. & Kimble, J. *Nature* **419**, 312–316 (2002).
- Richter, J. D. *Microbiol. Mol. Biol. Rev.* **63**, 446–456 (1999).
- Kadyk, L. C. & Kimble, J. *Development* **125**, 1803–1813 (1998).
- Aravind, L. & Koonin, E. V. *Nucleic Acids Res.* **27**, 1609–1618 (1999).
- Saitoh, S. *et al. Cell* **109**, 563–573 (2002).
- Read, R. L. *et al. Proc. Natl Acad. Sci. USA* (in the press).
- Castaño, I. B., Heath-Pagliuso, S., Sadoff, B. U., Fitzhugh, D. J. & Christman, M. F. *Nucleic Acids Res.* **24**, 2404–2410 (1996).
- Aphasizhev, R. *et al. Cell* **108**, 637–648 (2002).
- Martin, G., Keller, W. & Doublié, S. *EMBO J.* **19**, 4193–4203 (2000).

Stimulating illusory own-body perceptions

The part of the brain that can induce out-of-body experiences has been located.

Out-of-body' experiences (OBEs) are curious, usually brief sensations in which a person's consciousness seems to become detached from the body and take up a remote viewing position^{1–3}. Here we describe the repeated induction of this experience by focal electrical stimulation of the brain's right angular gyrus in a patient who was undergoing evaluation for epilepsy treatment. Stimulation at this site also elicited illusory transformations of the patient's arm and legs (complex somatosensory responses) and whole-body displacements (vestibular responses), indicating that out-of-body experiences may reflect a failure by the brain to integrate complex somatosensory and vestibular information^{1–3}.

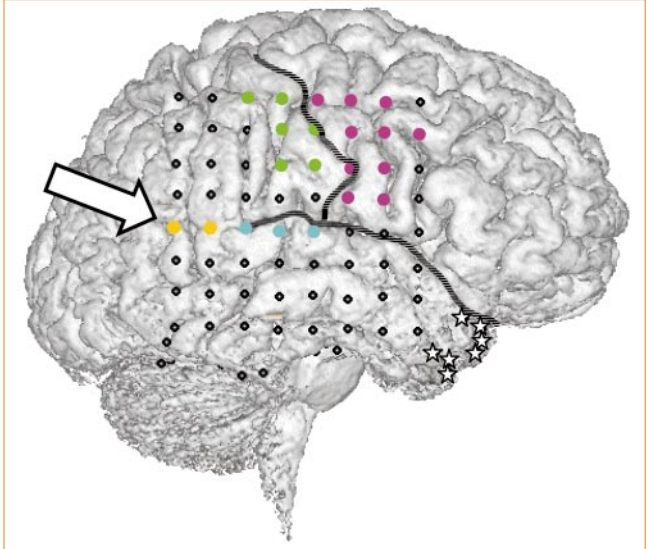
Our patient was a 43-year-old, right-handed woman who had suffered from complex partial seizures for 11 years; right temporal-lobe epilepsy was implicated. As magnetic-resonance imaging did not reveal any lesion, invasive monitoring was undertaken to localize the seizure focus precisely. Subdural electrodes were implanted to record seizures, and focal electrical stimulation was used to identify the vital cortex⁴.

Figure 1 shows the results of stimulation mapping and the electrode site on the right angular gyrus where stimulation repeatedly induced OBEs, as well as vestibular and complex somatosensory responses. Mapping of motor, somatosensory and auditory functions revealed no deviant brain pathology in this patient with respect to anatomical representations of cortical functions. The epileptic focus was located more than 5 cm anterior to the stimulation site, in the medial temporal lobe; electrical stimulation of this site did not induce OBEs, and these experiences were not part of the patient's habitual seizures.

Initial stimulations ($n = 3$; 2.0–3.0 mA) induced vestibular responses, in which the patient reported that she was "sinking into the bed" or "falling from a height". Increasing the current amplitude (3.5 mA) led to an OBE ("I see myself lying in bed, from above, but I only see my legs and lower trunk"). Two further stimulations induced the same sensation, which included an instantaneous feeling of "lightness" and "floating" about two metres above the bed, close to the ceiling.

The patient was then asked to watch her (real) legs during the electrical stimulation ($n = 2$; 4.0, 4.5 mA). As before, she was lying down (upper body supported at an angle of 45°, legs outstretched). This time, she reported seeing her legs "becoming shorter".

Figure 1 Three-dimensional surface reconstruction of the right hemisphere of the brain from magnetic-resonance imaging. Subdural electrodes were implanted in the brain of an epileptic patient undergoing presurgical evaluation; the locations at which focal electrical stimulation (ES) evoked behavioural responses are shown: magenta, motor; green, somatosensory cortex; turquoise, auditory cortex. Yellow, site at which out-of-body experience (OBE), body-part illusions and vestibular responses were induced (arrow). Stars indicate the epileptic focus in the medial temporal lobe. Informed consent was obtained from the patient and ES procedures conformed to the Declaration of Helsinki. Constant current (0.5–5.0 mA, 2-s train duration) was applied at 50 Hz in a bipolar manner through adjacent contacts⁴. Since undergoing a right anterior temporal lobectomy in 2000, the patient has been free of complex partial seizures.



If the patient's legs were bent before the stimulation (90° knee angle; $n = 2$; 4.0, 5.0 mA), she reported that her legs appeared to be moving quickly towards her face, and took evasive action.

When asked to look at her outstretched arms during the electrical stimulation ($n = 2$; 4.5, 5.0 mA), the patient felt as though her left arm was shortened; the right arm was unaffected. If both arms were in the same position but bent by 90° at the elbow, she felt that her left lower arm and hand were moving towards her face ($n = 2$; 4.5, 5.0 mA). When her eyes were shut, she felt that her upper body was moving towards her legs, which were stable ($n = 2$; 4.0, 5.0 mA).

These observations indicate that OBEs and complex somatosensory illusions can be artificially induced by electrical stimulation of the cortex. The association of these phenomena and their anatomical selectivity suggest that they have a common origin in body-related processing^{1–3}, an idea that is supported by the restriction of these visual experiences to the patient's own body.

During her OBE, the patient only 'saw' that part of her body that she also felt was modified during her body-transformation experiences. This contrasts with the non-corporeal visual hallucinations that are commonly induced by electrical stimulation at the parieto-temporal junction⁵. As suggested by previous neurological investigations on OBEs^{1–3} and other body-cognition disorders^{6–8}, the angular gyrus could be a

crucial node in a larger neural circuit that mediates complex own-body perception.

Out-of-body and body-transformation experiences are transitory and may disappear when a person attempts to inspect the illusory body or body part^{1–3}. Our findings suggest that changes in visual attention and/or current amplitude in the angular gyrus^{4,9} could bring about these phenomenological modifications.

Although we do not fully understand the neurological mechanism that causes OBEs, our results imply that vestibular processing² may be important. Although translational vestibular responses were evoked initially without an OBE and can be produced in isolation⁴, vestibular sensations of levitation and lightness^{1–3} accompanied OBEs in our patient. Also, the core region of the human vestibular cortex is situated close to the angular gyrus¹⁰. It is possible that the experience of dissociation of self from the body is a result of failure to integrate complex somatosensory and vestibular information.

Olaf Blanke*†, Stéphanie Ortigue†, Theodor Landis†, Margitta Seeck*

*Laboratory of Presurgical Epilepsy Evaluation, Program of Functional Neurology and Neurosurgery, University Hospitals of Geneva and Lausanne, Geneva 1211 and Lausanne 1011, Switzerland

e-mail: olaf.blanke@hcuge.ch

†Functional Brain Mapping Laboratory, Department of Neurology, Geneva University Hospital, 1211 Geneva, Switzerland

- Brugger, P., Regard, M. & Landis, T. *Cogn. Neuropsychiatr.* **2**, 19–38 (1997).
- Grüsser, O. J. & Landis, T. *Visual Agnosias and Other Disturbances of Visual Perception and Cognition* 297–303 (Macmillan, Amsterdam, 1991).
- Hécaen, H. & Ajuriaguerra, J. *Méconnaissances et Hallucinations Corporelles* 310–343 (Masson, Paris, 1952).
- Blanke, O., Perrig, S., Thut, G., Landis, T. & Seeck, M. *J. Neurol. Neurosurg. Psychiatr.* **69**, 553–556 (2000).
- Penfield, W. & Perot, P. *Brain* **86**, 595–696 (1963).
- Damasio, A. *The Feeling of What Happens: Body, Emotions and the Making of Consciousness* 213–215 (Vintage, London, 2000).
- Worthington, A. & Beevers, L. *Neurocase* **2**, 135–140 (1996).
- Halligan, P. W., Marshall, J. C. & Wade, D. T. *Cortex* **31**, 173–182 (1995).
- Nathan, S. S., Sinha, S. R., Gordon, B., Lesser, R. P. & Thakor, N. V. *Electroencephalogr. Clin. Neurophysiol.* **86**, 183–192 (1993).
- Lobel, E., Kleine, J., Leroy-Wilg, A., Le Bihan, D. & Berthoz, A. *J. Neurophysiol.* **80**, 2699–2709 (1998).

Competing financial interests: declared none.

Eukaryotic evolution

Early origin of canonical introns

Spliceosomal introns, one of the hallmarks of eukaryotic genomes, were thought to have originated late in evolution^{1,2} and were assumed not to exist in eukaryotes that diverged early — until the discovery of a single intron with an aberrant splice boundary in the primitive ‘protozoan’ *Giardia*³. Here we describe introns from a close relative of *Giardia*,

Carpediemonas membranifera, that have boundary sequences of the normal eukaryotic type, indicating that canonical introns are likely to have arisen very early in eukaryotic evolution.

Carpediemonas membranifera is a poorly studied, free-living microbial eukaryote that is considered to be a relative of *Giardia* on the basis of its morphology⁴. Using the polymerase chain reaction (PCR) with *Carpediemonas* genomic DNA as template, we determined the partial sequences of two distinct carbamate kinase genes from this organism. In both genes, an insertion of 33 or 31 nucleotides interrupts the similar protein-coding sequence shared with carbamate kinase genes from other organisms (Fig. 1a). These insertions are bounded by guanine and thymine (GT) nucleotides at the 5’ end and adenine and guanine (AG) nucleotides at the 3’ end, which is a characteristic of most of the spliceosomal introns that interrupt protein-coding genes in other eukaryotes.

We used PCR with reverse transcription to recover the messenger RNA sequence of one of the two *Carpediemonas* carbamate kinase genes. This sequence lacks the insertion, which is presumably removed (spliced) from the messenger RNA before translation. We conclude that the insertions in the *Carpediemonas* carbamate kinase genes are canonical ‘GT...AG’ spliceosomal introns, albeit comparatively small ones.

To determine the evolutionary affinities of *Carpediemonas*, we used PCR to amplify near-complete sequences for two genes that encode cytosolic heat-shock protein 70 (Hsp70). We also sequenced a cloned Hsp70 gene from *Spironucleus barkhanus*, a very close relative of *Giardia*. Maximum likelihood analysis of Hsp70 proteins reveals a specific evolutionary relation between *Carpediemonas*, *Giardia* and *Spironucleus* (Fig. 1b); three other molecular markers also support this relationship⁵.

The single intron found in a *Giardia* gene has a non-canonical CT dinucleotide at its 5’ splicing boundary³, which could be interpreted as a ‘frozen’ primitive eukaryotic condition: canonical ‘GT...AG’

spliceosomal introns might then be a later innovation in more modern cells. Our results indicate that this is not the case, however, as canonical introns seem to be an ancestral feature of the larger evolutionary grouping that includes *Giardia* and *Carpediemonas*. The aberrant *Giardia* intron probably represents a lineage-specific (or intron-specific) secondary alteration of the 5’ splice boundary.

The extremely early divergence attributed to *Giardia* is based on the absence or aberration of many typical eukaryotic features, such as mitochondria and introns, and on its arguably deep-branching position in many phylogenetic trees^{6–9}. The grouping of *Giardia* with *Carpediemonas* (which, as well as canonical introns, has organelles that are probably derived from mitochondria⁴) weakens this argument for early divergence.

Irrespective of the true evolutionary position of *Giardia*, the only potentially ‘early’ eukaryotic group in which introns have not been found are the parabasalids, such as *Trichomonas*^{10,11}. *Trichomonas* is already known to possess some of the cellular machinery for intron splicing¹², however, and there is evidence to indicate that it is evolutionarily affiliated with *Giardia* and its relatives^{7,13} (and is slightly misplaced in many phylogenies, including that shown in Fig. 1b). An affiliation with *Giardia* implies a similar closeness to *Carpediemonas*, and it is likely that parabasalids have, or had, canonical introns. There is now every reason to assume that canonical introns were present in the most recent common ancestor of living eukaryotes.

Alastair G. B. Simpson, Erin K. MacQuarrie, Andrew J. Roger

Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada e-mail: simpson@hades.biochem.dal.ca

- Palmer, J. D. & Logsdon, J. M. *Curr. Opin. Genet. Dev.* **1**, 470–477 (1991).
- Logsdon, J. M. *Curr. Opin. Genet. Dev.* **8**, 637–648 (1998).
- Nixon, J. E. *et al. Proc. Natl Acad. Sci. USA* **99**, 3701–3705 (2002).
- Simpson, A. G. B. & Patterson, D. J. *Eur. J. Protistol.* **35**, 353–370 (1999).
- Simpson, A. G. B. *et al. Mol. Biol. Evol.* **19**, 1782–1791 (2002).
- Cavalier-Smith, T. *Trends Genet.* **7**, 145–148 (1991).
- Embley, T. M. & Hirt, R. P. *Curr. Opin. Genet. Dev.* **8**, 624–629 (1998).
- Roger, A. *J. Am. Nat.* **154** (suppl.), 146–163 (1999).
- Sogin, M. L. *Curr. Opin. Genet. Dev.* **7**, 792–799 (1997).
- Johnson, P. J. *Proc. Natl Acad. Sci. USA* **99**, 3359–3361 (2002).
- Archibald, J. M., O’Kelly, C. J. & Doolittle, W. F. *Mol. Biol. Evol.* **19**, 422–431 (2002).
- Fast, N. M., Logsdon, J. M. & Doolittle, W. F. *Mol. Biochem. Parasitol.* **99**, 514–522 (1999).
- Dacks, J. B. & Roger, A. J. *J. Mol. Evol.* **48**, 779–783 (1999).

Competing financial interests: declared none.

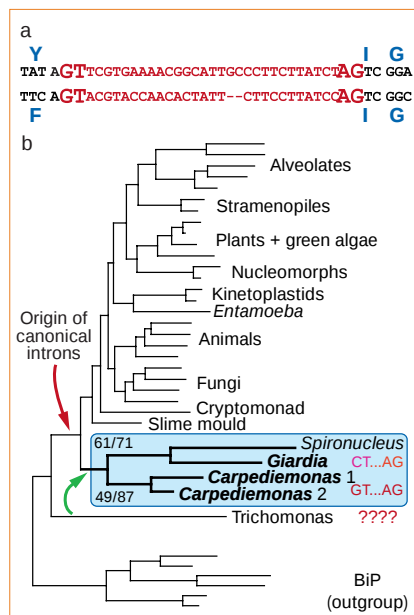


Figure 1 Introns and evolutionary affinities of *Carpediemonas*. **a**, Portions of two *Carpediemonas* carbamate kinase genes, showing intron sequences (red) interrupting the protein-coding sequence (in blue). The introns have canonical splice boundaries (GT...AG; large red type). **b**, Maximum-likelihood evolutionary tree of eukaryotic cytosolic Hsp70 proteins ('T + invariable sites' model). Endoplasmic-reticulum Hsp70 ('BiP') is used as an outgroup. The grouping of *Carpediemonas* with *Giardia* and *Spironucleus* is shown in the blue box; statistical support (bootstrap percentages) for this grouping is assessed using likelihood (upper left of box), and likelihood distance (lower left). The higher percentages (right in each pair) apply when the outgroup is omitted. The basal placement of *Trichomonas* is weakly supported with likelihood (21%); green arrow shows a more plausible position on the basis of other evidence^{7,13}. The intron splice boundaries for the relevant groups and the origin of canonical introns are shown in red. New sequences have been deposited at GenBank under accession numbers AY131204–AY131209.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com

Structure of the Sec23/24–Sar1 pre-budding complex of the COPII vesicle coat

Xiping Bi, Richard A. Corpina & Jonathan Goldberg

Howard Hughes Medical Institute and the Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

COPII-coated vesicles form on the endoplasmic reticulum by the stepwise recruitment of three cytosolic components: Sar1–GTP to initiate coat formation, Sec23/24 heterodimer to select SNARE and cargo molecules, and Sec13/31 to induce coat polymerization and membrane deformation. Crystallographic analysis of the *Saccharomyces cerevisiae* Sec23/24–Sar1 complex reveals a bow-tie-shaped structure, 15 nm long, with a membrane-proximal surface that is concave and positively charged to conform to the size and acidic-phospholipid composition of the COPII vesicle. Sec23 and Sar1 form a continuous surface stabilized by a non-hydrolysable GTP analogue, and Sar1 has rearranged from the GDP conformation to expose amino-terminal residues that will probably embed in the bilayer. The GTPase-activating protein (GAP) activity of Sec23 involves an arginine side chain inserted into the Sar1 active site. These observations establish the structural basis for GTP-dependent recruitment of a vesicular coat complex, and for uncoating through coat-controlled GTP hydrolysis.

Recent research on vesicle transport pathways responsible for the growth and maintenance of organelles has shown that vesicle budding and fusion mechanisms are conserved in organisms from yeast to humans¹. Budding occurs by the action of cytoplasmic coat protein complexes (COPs) that polymerize on the membrane surface, capturing cargo and SNARE molecules in the process, and deforming the membrane to sculpt out vesicles². Small GTP-binding proteins of the ARF family regulate coat assembly and disassembly by providing both the spatial cue—activated ARF–GTP—that restricts coat recruitment to specific export sites^{3–5}, and the temporal cue—controlled GTP hydrolysis—that triggers uncoating^{6–8}. Cells contain a variety of COPs, including COPI or coatomer, COPII, and numerous isoforms of clathrin/adaptin complexes², each COP specifying a discrete transport step according to its affinities for a particular export site and for a particular set of SNARE molecules gathered up during budding to programme the vesicle for fusion with its target membrane^{9–11}.

COPII-coated vesicles bud from the endoplasmic reticulum (ER) to export newly synthesized proteins to the Golgi complex. The components of COPII comprise the ARF-related molecule Sar1 and two large heterodimeric complexes, Sec23/24 and Sec13/31 (ref. 5). These five proteins when supplied with GTP can bud characteristic 50–90-nm vesicles from synthetic liposomes, provided these contain about 10% acidic phospholipids for effective binding of Sec23/24 (ref. 12). *In vivo*, budding is initiated by the exchange of GDP for GTP on Sar1 catalysed by Sec12, an ER-localized guanine-nucleotide exchange factor, or GEF^{13–15}. Sar1–GTP then binds to the membrane, probably by embedding an N-terminal α -helix¹⁶ in the bilayer in the manner delineated for ARF1 (refs 17–20). Sar1–GTP recruits Sec23/24 through an interaction with the Sec23 subunit⁷, and this pre-budding complex binds Sec13/31, which is believed to act, like clathrin, to polymerize the coat and deform the membrane into a bud^{5,12}. Electron microscopy (EM) of isolated complexes shows that Sec23/24 resembles a bow tie and that Sec13/31 is elongated and apparently rather flexible²¹. EM analysis of the COPII vesicle coat suggests a layered structure comprising an inner shell of Sec23/24 and an outer shell of Sec13/31 (ref. 22), recapitulating the sequential binding mechanism observed in budding experiments.

Studies of the cargo-selection function of Sec23/24–Sar1 in yeast and mammalian systems have shown that ER-to-Golgi SNAREs and at least some integral membrane cargo proteins can bind directly to the pre-budding complex^{10,23–26}, leading to their concentration in COPII vesicles^{27,28}. A final function of Sec23/24—one that highlights the dynamic nature of vesiculation—is as the GAP for the otherwise inert GTPase activity of Sar1–GTP⁷. It seems that, as the pre-budding complex forms, a slow GTP hydrolysis reaction is programmed into the assembling coat to limit its lifetime and to couple coating to uncoating^{7,8}.

In the present study, molecular aspects of COPII function in ER export were addressed by determining the structure of the Sec23/24–Sar1 complex. Its domain composition and architecture bear no resemblance to AP2 adaptin or clathrin^{29,30}. An extensive membrane-interaction area is formed by all three proteins and is concave and basic to match the COPII vesicle surface. The nature of the interface between Sec23 and Sar1 explains how GTP binding initiates coat recruitment and how coat recruitment initiates GTPase activity, suggesting a molecular-level picture of the control of coat assembly and disassembly by a cycle of GTP binding and hydrolysis.

Structure determination

All structural studies used full-length recombinant yeast Sec23 and Sec24 proteins. Because full-length Sar1–GTP tends to aggregate in solution³¹, a truncated form was used that lacks the N-terminal 23 amino acids. Biochemical studies show that ARF proteins truncated in this way specifically lose the membrane-interaction function and retain all protein-interaction functions^{32,33}; structural studies of ARF proteins in truncated and full-length forms show that N-terminal truncation has no evident effect on conformation^{19,34,35}.

The atomic model of Sec23/24–Sar1 was built from structures of Sec23/24 and Sec23–Sar1 determined by X-ray crystallography. Crystals of the yeast Sec23–Sar1 complex (relative molecular mass (M_r) of Sec23 85K, Sar1 19K) bound to the non-hydrolysable GTP analogue 5'-guanylyl imidodiphosphate (GppNHp) and Mg^{2+} grew in the space group $P2_12_12_1$ with two complexes in the asymmetric unit. The initial electron density map was calculated to 2.7 Å resolution from a multiwavelength anomalous diffraction

(MAD) data set incorporating selenium as the anomalous scatterer. The structure was refined with native X-ray data to 2.5 Å resolution (Supplementary Information Table 1). The yeast Sec23/24 complex (Sec24 M_r 104K) crystallized in the space group $P2_12_12_1$. This structure was determined by the molecular replacement method, and refined with data to 2.75 Å resolution (Supplementary Information Table 1).

Comparison of the two crystal structures shows that only minor changes occur in Sec23 upon interaction with either Sar1 or Sec24. In addition, the binding sites on Sec23 for the two proteins are separated by more than 25 Å, and hence it was possible to construct a straightforward composite model of Sec23/24–Sar1 from the two experimental structures.

Architecture and membrane interaction

Figure 1a shows the inner face of the coat that will probably appose the membrane, and highlights the bow-tie shape of the complex. The folds of Sec23 and Sec24 are closely related, and the two subunits interact about a dyad (Fig. 1a is viewed along this axis). Comparison with the 'side' view (Fig. 1b) shows that Sec23/24–Sar1 is elongated in directions parallel to the membrane so as to cover an extensive membrane area; it measures ~150 Å in the long direction, and varies along the bow tie from ~75 Å at the Sec24 end, to as little as 30 Å at the dimer interface, to ~100 Å at the Sec23–Sar1 end (Fig. 1a). By contrast, the complex is relatively thin and extends uniformly only about 40 Å in the direction away from the membrane (Fig. 1b), except for two protrusions formed by a gelsolin-like domain on each subunit that extend ~60 Å (Figs 2a and 3a). For comparison, the thickness of the COPII coat is ~100 Å, on the basis of EM images²².

A distinctive feature of the pre-budding complex is its concave

inner surface, which seems to match the size of a 60-nm vesicle (Fig. 1b and central images in Fig. 3). Although Sec13/31 binding is thought to provide the driving force for vesiculation, the observed curvature suggests that Sec23/24–Sar1 will favour membrane deformation and perhaps exert some influence on the final dimensions and rather uniform size of COPII vesicles that bud from purified ER membranes⁵. Consistent with this idea, *in vitro* experiments show that the yeast Sec24 homologue Lst1 can combine with Sec23 to form mixed Sec23/24–Sec23/Lst1 coats that bud somewhat larger (~12 nm greater in diameter) and size-heterogeneous vesicles³⁶.

Sec23/24–Sar1 appears complementary to the vesicle surface both in shape and chemistry. Examination of the protein electrostatic surface reveals that the concave, inner face is generally positively charged, and the outer face much more acidic (Fig. 3a). This polarization is conserved from yeast to human Sec23/24 complexes (as a general conservation of the surface regions rather than invariance of numerous arginine and lysine side chains). These observations can explain the results of *in vitro* budding experiments that demonstrated a requirement for acidic phospholipids for the Sar1-dependent binding of Sec23/24 to synthetic liposomes¹².

The Sec23/24–Sar1 structure reveals three determinants of membrane interaction: the curved form and positively charged surface as described above, plus the orientation of the Sar1 N terminus within the complex. Sar1–GppNHp interacts extensively with Sec23 to form a continuous surface that extends the membrane-interaction area (Figs 1a and 2a), and this surface includes the Sar1 N terminus (Gly24 is the first modelled residue), which projects towards the membrane, probably to embed N-terminal residues in the bilayer (see below). If Sec23/24–Sar1 interacts closely with the membrane as these observations suggest, then the concave inner surface of one complex would cover ~8,200 Å² of membrane, roughly 0.7% of the

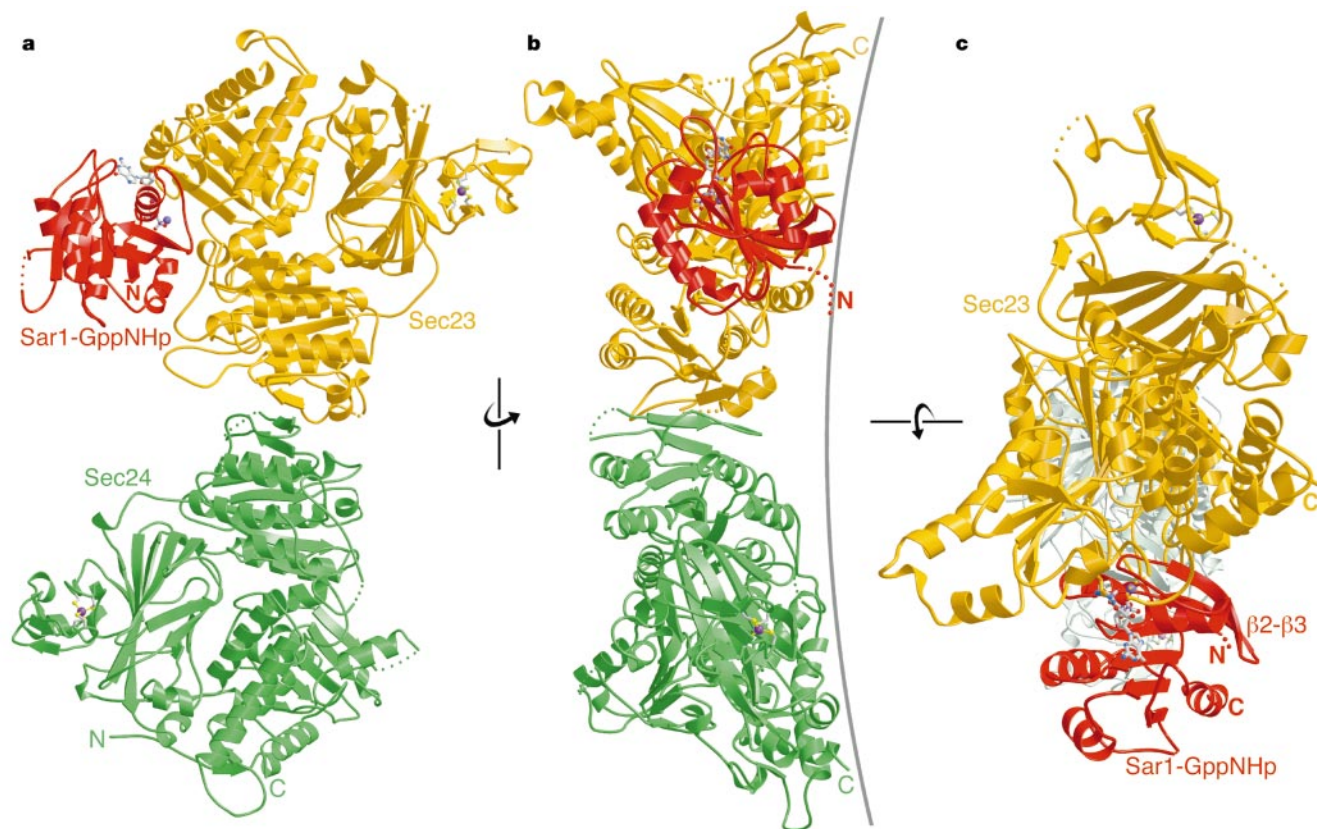


Figure 1 Structure of the Sec23/24–Sar1 pre-budding complex. Ribbon representation is shown as successive 90° rotations. Sec23 is yellow, Sec24 green and Sar1 red. **a**, Front view along the dyad relating the coat subunits, with the membrane-proximal surface

facing forward. **b**, Side view. The grey line indicates the curvature of a 60-nm COPII vesicle, drawn to scale. **c**, Top view. The three regions of Sar1 that would face the membrane are labelled C, N and β2–β3.

surface area of a 60-nm vesicle; and 60 such complexes would provide ~45% coverage of the vesicle surface. It will be interesting to compare this estimate with EM analyses of the lattice of subunits formed in the assembled coat. But current EM images of the COPII coat show rather densely packed features²² and lack the contrast of those produced in clathrin studies, which benefit from the open network of that lattice^{2,22}.

Domain organization of Sec23 and Sec24

The structural relatedness of the subunits was suggested in EM images showing the bow-tie complex as two heart-shaped halves that interact about a dyad²¹. The crystal structure reveals that the folds of the subunits are closely related, but that the sequences have diverged substantially so that the residual sequence identity is only ~14% according to a structure-based alignment (Fig. 2b), and alignments calculated using ClustalW³⁷ alone give somewhat

spurious results. The alignment in Fig. 2b includes homologous regions of ~750 residues, and excludes the divergent 180 residues at the Sec24 N terminus, the initial 132 of which are disordered in the crystals, consistent with the low content of hydrophobic side chains and high content of glutamine and proline residues in this region.

The Sec23 and Sec24 polypeptides fold into five distinct domains (Fig. 2): a β -barrel, a zinc finger, an α/β vWA³⁸ or 'trunk' domain, an all-helical region and a carboxy-terminal domain that closely resembles a gelsolin module³⁹. The β -barrel is formed from ~180 residues contributed by three polypeptide segments (see Fig. 2b). The strands of the barrel lie roughly parallel to the membrane such that one end of the barrel forms part of the concave inner surface of the coat and the other end part of the membrane-distal surface. The barrel is constructed from two apposed sheets: the six-stranded sheet β 4-10-24-23-19-21 (numbered in the membrane distal-to-

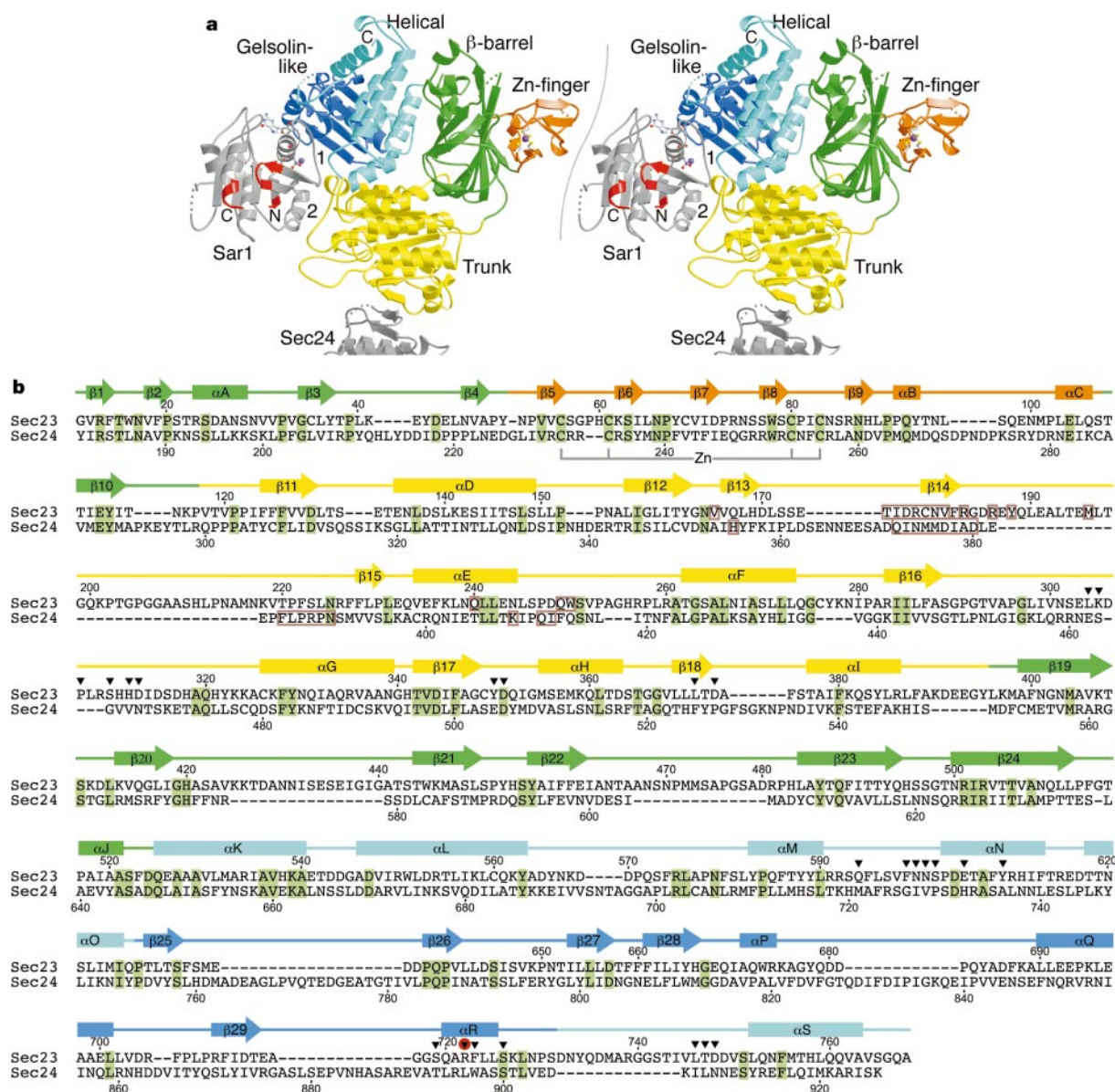


Figure 2 Domain structure and relatedness of Sec23 and Sec24. **a**, Stereo view, coloured according to the domain structure of Sec23, and oriented as in Fig. 1a. Sar1 and Sec24 are grey. The switch 1 and 2 regions of Sar1 are labelled 1 and 2, and the three membrane-proximal regions are red. **b**, Structure-based sequence alignment of *Saccharomyces cerevisiae* Sec23 and Sec24. The consensus secondary structure is

shown above the sequences, coloured as in **a**, with α -helices depicted as cylinders and β -strands as arrows. Residues making intersubunit (between Sec23 and Sec24) interactions are outlined in brown, those residues of Sec23 contacting Sar1–GppNHp are denoted with black arrowheads and a red sphere for the catalytic arginine residue 722, and sequence identity is highlighted in green.

proximal direction) faces partly towards the zinc finger and partly towards solvent; sheet β 1-3-22-20 faces the helical domain.

The ~ 55 -residue zinc-finger domains lie against the β -barrels at the periphery of the complex (Figs 1a and 2a). The Zn-Cys₄ cluster is situated towards the membrane-proximal surface and several neighbouring arginine and lysine side chains on Sec23 and particularly Sec24 contribute to the basic inner surface (Fig. 3a). The trunk domain is formed from a single ~ 250 -residue segment plugged into the barrel between strands β 1 and β 19. The trunk has an α/β fold with the vWA-type topology³⁸ and it forms the dimer interface, primarily involving strand β 14 on Sec23 and Sec24 (Fig. 2); in addition, the trunk domain of Sec23 contacts Sar1. The α -helical domain forms from an ~ 105 -residue segment with the C-terminal ~ 30 residues. Helices α K-L-M form a three-helix bundle with α L facing the concave inner surface. Three additional helices (N, O and S) complete the domain, with the α M- α N linker contacting Sar1. Finally, the gelsolin-like domain is formed by a contiguous stretch of 105 residues in Sec23, and a 150-residue region in Sec24 that is longer due to three insertions, all of which are disordered in the crystals. This is the only domain that does not contribute to the inner surface of the coat; instead it lies against the helical domain, forms critical catalytic interactions with Sar1 and the nucleotide, and extends radially away from the membrane, perhaps to form an interaction site for Sec13/31 (Figs 2a and 3a).

Heterodimer interface and Sec24 isoforms

In addition to the essential Sec23 and Sec24 proteins, yeast encode two (and mammals at least four) non-essential Sec24 homologues, Iss1 and Lst1, which are 55 and 23% identical to the Sec24 polypeptide; both form heterodimers with Sec23 (refs 36, 40). All three Sec24 isoforms interact with the ER-to-Golgi SNARE Sed5/

syntxin-5 (ref. 40), but only the Sec23/Lst1 pairing is proficient at packaging the plasma membrane ATPase Pma1 into vesicles³⁶. Defining the interactions between these pre-budding complexes and sorting signals on SNARE and cargo molecules is a challenge for future studies, but at this stage an analysis of the conserved surface of Sec24 can suggest potential binding sites. In particular, the solvent-facing sheet of the β -barrel exposes a patch of highly conserved and hydrophobic side chains: Tyr 296, Val 557 and Leu 616 (Fig. 3b).

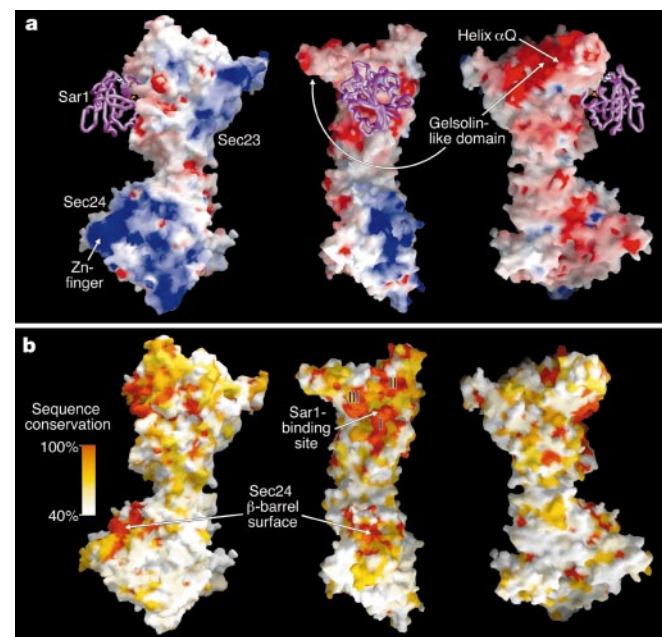


Figure 3 Surface features of the Sec23/24 complex **a**, Molecular surface coloured according to electrostatic potential⁵⁰: negative potential is red and positive potential blue. Sar1 is drawn as a backbone tube in magenta. The three views are successive 90° rotations around a vertical axis. In the left image, the membrane-proximal surface faces forwards. **b**, Surface of Sec23/24 coloured according to sequence conservation of the underlying residues in an alignment of Sec23 sequences from six organisms (human Sec23A, *Drosophila melanogaster*, *Neurospora crassa*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana* and *Saccharomyces cerevisiae*) and Sec24 sequences from four organisms (human Sec24A, *S. pombe*, *A. thaliana* and *S. cerevisiae*). Labels I, II and III indicate three highly conserved patches (see text for details). The orientations are the same as in **a**, with Sar1 omitted.

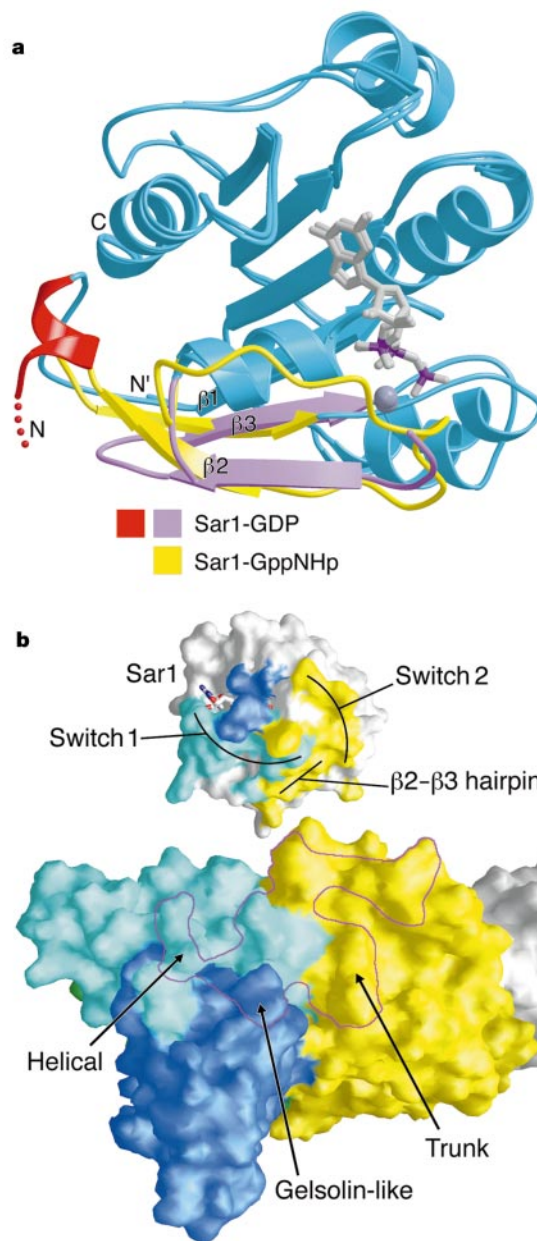


Figure 4 Conformational switching and protein-protein interactions. **a**, The nature of the Sar1 GDP-GTP conformational switch is suggested by superimposing mammalian Sar1-GDP¹⁶ and yeast Sar1-GppNHp (from the Sec23-Sar1 complex). GTP-induced translation of β -strands 2 and 3 from the diphosphate (pink) to triphosphate (yellow) conformation eliminates the binding site for the N-terminal region (red) thought to interact with lipid membrane. Dark grey spheres are magnesium ions. N' denotes the amino terminus (Gly 24) of the truncated Sar1 protein used in this study. **b**, 'Open book' view of the interfacial surfaces, with Sec23 and Sar1 rotated apart by 140°. The molecular surface of Sar1 is coloured to indicate the footprint of Sec23. The corresponding contact region on Sec23 is outlined in magenta. Colouring as in Fig. 2.

Sec23 recognizes Sec24 through an interface that buries $\sim 900 \text{ \AA}^2$ of Sec24 surface area. The contributing residues, chiefly from strand $\beta 14$ and the $\beta 14$ – $\beta 15$ loop, are not highly conserved among the yeast Sec24 isoforms, perhaps because strand $\beta 14$ engages primarily in main-chain hydrogen bonds with the symmetry-related strand of Sec23, to create an inter-subunit β sheet joining the trunk domains (Figs 1 and 2). The exception is the $\beta 14$ – $\beta 15$ loop region FLP (positions 385–387), which is conserved among the yeast Sec24 homologues and across species. Residues Phe 385 and Pro 387 from this region form van der Waals interactions with the universally conserved residues VFR (181–183) of Sec23 (Fig. 2b).

Sar1 conformational switch

Studies primarily on mammalian ARF1 have shown that ARF-family proteins possess a ‘membrane anchor’ that can be retracted in the GDP-bound state and exposed in the GTP state^{3,18,20}. The anchor comprises an N-terminal amphipathic α -helix^{17,34} that forms tight interactions with the membrane by inserting hydrophobic side chains into the bilayer¹⁸. In addition, the anchor on ARF proteins (although not on Sar1) is N-myristoylated, apparently not for the purpose of enhancing the membrane affinity of the GTP state but rather to increase the encounter frequency of ARF–GDP and membranes^{41,42}.

Biochemical knowledge of Sar1 is less advanced than that of ARF1, but the extent of sequence homology between these proteins (yeast Sar1 and human ARF1 are 33% identical) implies a common switching mechanism. And this is supported by the marked structural similarity of ARF1–GDP and Sar1–GDP, and of ARF1–GppNHp and Sar1–GppNHp (refs 16, 17, 19 and this study). Comparison of Sar1 in its GDP and GTP states provides a detailed view of the altered bonding around the γ phosphate that underlies the conformational switch, and this seems to be essentially the same as observed for ARF1, ARF6 and ARL2 (refs 19, 34, 35). For example, residue Thr 54 in the switch 1 region of Sar1–GppNHp (equivalent to Thr 35 in Ras) forms bonds to the γ phosphate and Mg^{2+} , and residue Gly 76 (Gly 60 in Ras) in switch 2 coordinates the γ phosphate.

Figure 4a shows the conformational change driven by these and associated interactions, and indicates how GTP binding may trigger the exposure of the Sar1 membrane anchor (red helix in Fig. 4a). In Sar1 proteins, an N-terminal extension of about ten residues

replaces the myristoyl group and precedes the amphipathic α -helix. In Sar1–GDP, the helix is retracted into a surface pocket formed in part by the linker region of the $\beta 2$ – $\beta 3$ hairpin¹⁶. The structure of Sar1–GppNHp shows that GTP binding causes an approximately 7 $\text{\AA}$ displacement of the hairpin—involving a change in register of the hydrogen bonding between strands $\beta 3$ and $\beta 1$ —to eliminate the pocket for the anchor (Fig. 4a). This seems to explain how cytosol/membrane translocation of the protein can be controlled according to the type of nucleotide bound at its active site. And on the basis of these and analogous observations in the ARF1, ARF6 and ARL2 systems, it is probable that this mechanism is used for all membrane-dependent transactions in cells involving ARF-family proteins^{19,34,35}.

GTP-dependent coat recruitment

The arrangement of Sar1 in the pre-budding complex establishes how an activated ARF-family protein is oriented with respect to the membrane surface (Figs 1 and 2a). Figures 1b and 2a indicate that the β strands of Sar1 ($\beta 1$ – $\beta 3$ in particular) are approximately parallel to the membrane normal, such that three regions of Sar1 juxtapose the membrane: the anchor and additional N-terminal residues (in sum, residues 1–23, which are absent from the crystal structure), the linker region of the $\beta 2$ – $\beta 3$ hairpin (residues 63–68), and C-terminal residues 187–190. Two basic residues, Lys 22 and Lys 68, which may contribute to membrane interaction, are located in these regions. More directly, *in vitro* studies on ARF1 have pinpointed interactions between acidic phospholipids and basic side chains in these regions, specifically the N-terminal residues Lys 15 and Lys 16, and C-terminal residues Arg 178 and Lys 181 (ref. 43; also note ARF1 residue Lys 59 in the $\beta 2$ – $\beta 3$ linker).

The binding of GTP to Sar1 has two conformational consequences—the anchor is released for membrane interaction and the affinity for Sec23/24 is enhanced—the combined effect of which is membrane recruitment and formation of the pre-budding complex. Sec23 forms an extensive interface with Sar1–GppNHp (Fig. 4b) that encompasses $\sim 1500 \text{ \AA}^2$ of surface area, roughly 20% of the total Sar1 surface. Comparison with uncomplexed Sec23 shows that recruitment to Sar1 causes a subtle rotation of the barrel, zinc-finger, helical and gelsolin-like domains together as a rigid body with respect to the trunk domain, such that the curvature of Sec23/24 increases slightly, by about 1.5° . Sec23 recognizes the GTP state

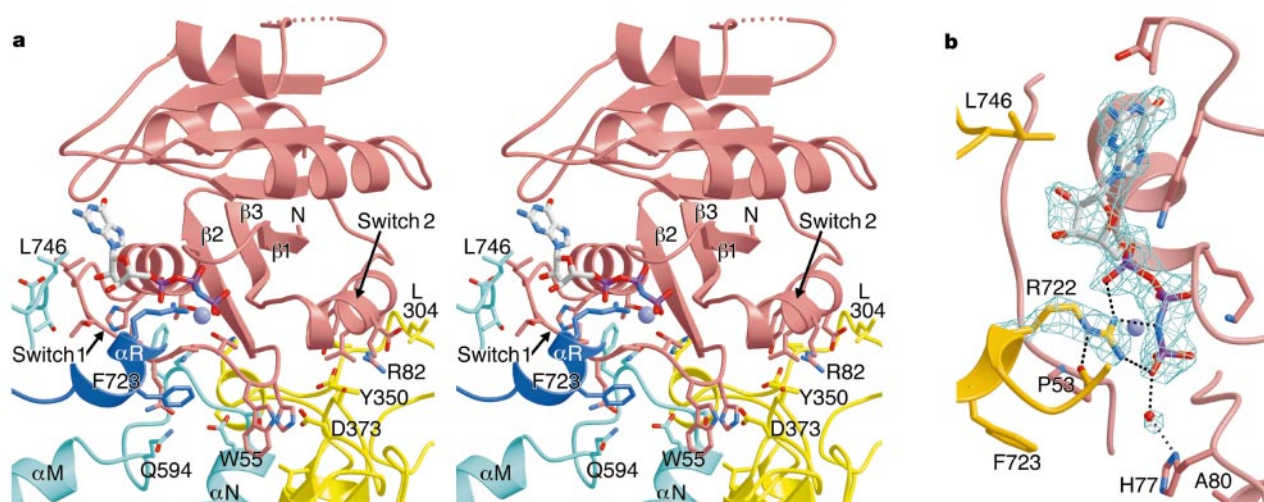


Figure 5 GTP-dependent coat recruitment and coat-dependent GTP hydrolysis **a**, Stereo view showing interactions between Sar1–GppNHp (pink) and Sec23 (coloured as in Fig. 2). Six amino-acid residues in Sec23 are labelled (triple-digit numbers), one from each discrete structural element that contributes to the interface. Sar1 residues W55 in switch 1 and R82 in switch 2 are labelled. Oxygen atoms are red, nitrogen atoms are blue,

phosphorus atoms are magenta, and the magnesium ion is a blue-grey sphere. For clarity, portions of Sec23 have been removed from the foreground. **b**, Difference electron density (SA omit⁴⁹) map at 2.5 $\text{\AA}$ resolution showing GppNHp, the catalytic residue Arg 722 of Sec23, and a water molecule (red sphere) positioned possibly for nucleophilic attack. Sar1 is pink and Sec23 is orange.

by contacting the switch 1, switch 2 and $\beta 2$ – $\beta 3$ hairpin regions of Sar1 using residues from loops rather than secondary structure elements. Switch 1 is contacted primarily by the helical domain, and switch 2 and the hairpin by loops from the trunk domain that emanate from the carboxy ends of the β strands (Figs 2b, 4b and 5a). (Switch 1 is defined as Sar1 residues Lys 44 to Pro 57, and switch II as Asp 73 to Pro 89.)

The Sar1-binding site is the most conserved region of extended surface on Sec23/24 (Fig. 3b), and of the 24 residues contacting Sar1–GppNHP (calculated with a 3.8 Å cutoff), 12 are invariant among Sec23 sequences in the six organisms analysed (see legend to Fig. 3). These 12 cluster in three regions—labelled I, II and III in Fig 3b—of which patch I, formed by the αM – αN loop of the helical domain (Figs 2b and 5a), seems most critical for recognition of Sar1–GTP. The invariant residues FNNS (599–602) and Glu 605 of this loop contact switch 1 and the preceding α -helix (Fig. 5a). First, residue Phe 599 plugs into a hydrophobic pocket created by switch 1, and Asn 600 accepts a hydrogen bond from the side-chain OH of Thr 54, which also engages Mg^{2+} (but only in the GTP state). Second, Ser 602 and Glu 605 form interactions with switch 1 residues WHP (55–57), and these are reinforced by contacts involving trunk-domain residues from loops $\beta 17$ – αH and $\beta 18$ – αI (Figs 2b and 5a). The conservation of His 56 and Pro 57 in Sar1 proteins, and the systematic change from WHP (55–57) to IGF (49–51) in mammalian ARF1 and related isoforms suggests that this set of interactions may have a role in Sec23 discrimination by Sar1 and ARF proteins.

The interaction patch II (Fig. 3b) is formed by the αR – αS loop of the helical domain, which contacts residues RLA (47–49) of switch 1, and is notable owing to a van der Waals interaction between Leu 746 and the guanine base (Fig. 5). In general, the Sec23–Sar1 interface reveals a combination of hydrophobic and polar interactions, formed predominantly by side-chain–side-chain contacts, with few main-chain atoms involved. The final set of interactions, involving patch III, is formed between the gelsolin-like domain and the nucleotide-binding site of Sar1, and this may function not to recruit Sec23/24 in response to GTP binding but to control GTP hydrolysis and uncoating.

GTP hydrolysis and coat disassembly

Sar1–GTP, like ARF1–GTP, is catalytically inert. The problem of how GTP hydrolysis is triggered to cause COPII coat disassembly was resolved by the discovery that Sec23 is the Sar1 GAP⁷, but this in turn raised the question of how the rate of COPII-mediated budding competes with the rate of COPII-mediated disassembly⁸. A ‘two gear’ mechanism has been proposed for the control of GTP hydrolysis, on the basis of kinetic studies showing that the lifetime of the Sec23/24–Sar1–GTP complex is relatively long (~ 30 s), and hydrolysis is accelerated about tenfold after Sec13/31 recruitment⁸. Hence, the COPII coat triggers its own disassembly, but the lifetime of GTP on Sar1 may be matched to the rate of budding to allow pre-budding complexes time to gather SNARE and cargo molecules, and completion of polymerization before disassembly⁸.

Inspection of the Sec23–Sar1 interface reveals the mechanism of GTPase stimulation (Fig. 5). Sec23 inserts an arginine side chain, Arg 722, into the Sar1 active site to form bonds to the phosphates through its guanidinium group that will neutralize negative charges in the transition state. This type of mechanism, involving an arginine residue (arginine ‘finger’) supplied *in trans* to a G-protein active site, is a feature of many Ras-family proteins⁴⁴. Evidently, isolated Sar1–GTP is inactive in part because it lacks the requisite catalytic machinery. Arg 722 and adjoining residues form helix αR on one face of the gelsolin-like domain (invariant patch III in Fig. 3b), and side-chain atoms of Ser 719, Arg 722, Phe 723 and Ser 726 directly contact Sar1. The side-chain OH of Ser 719 bonds to Asp 32 of the P-loop, but more importantly its C’O group forms a hydrogen bond to the catalytic Arg 722 to assist its positioning

(Fig. 5b). Likewise, Phe 723 of Sec23 stabilizes the productive configuration at the active site through hydrophobic bonds to Pro 53 of switch 1, creating a sandwich of van der Waals contacts, the layers of which comprise Phe 723–Pro 53–Arg 722–P-loop (Fig. 5). This rigid arrangement may explain why Arg 722 forms a productive complex with the phosphates even in this ground-state structure, unlike in the Ras- and Rho-GAP systems where the arginine finger engages phosphates only in the transition state⁴⁴.

What can be inferred about the lack of GTPase activity of isolated Sar1–GTP? Studies on other GAP systems have shown that GAPs act in two ways to cause rate acceleration: by inserting an arginine residue, and by stabilizing switch 2 primarily to position a catalytic glutamine side chain (Gln 61 in Ras) and its associated nucleophilic water molecule⁴⁴. In Sar1 proteins a histidine residue (His 77) replaces the glutamine, and in electron density maps His 77 bonds to an appropriately positioned water molecule (Fig. 5b). However, the position of switch 2 at the Sec23–Sar1 interface is unchanged from the position observed for the corresponding element in uncomplexed ARF1–GppNHP¹⁹, suggesting that reconfiguration of switch 2 is not a principal feature of Sec23 action. Additional kinetic/mutagenetic studies are needed to assess what, if any, catalytic changes are caused by Sec23 in addition to Arg 722 insertion.

The lack of movement in switch 2 is at first sight surprising because the position it adopts is thought to contribute to ARF1 inactivity through misalignment of the catalytic glutamine, Gln 71 (ref. 19). Comparative analysis of the Sec23–Sar1, ARF1 and Ras–RasGAP active sites suggests that Sar1 proteins may avoid alterations to switch 2 by bonding the nucleophilic water in a new way. The uncommon substitution of the catalytic glutamine by histidine, and the change from QXXI (71–74) in ARF1 to HXXA (77–80) in Sar1 allows His 77 to lie against Ala 80 (Fig. 5b) and bond the nucleophilic water molecule in a manner not seen in other G-protein structures, through a configuration unfeasible with the bulkier isoleucine and longer glutamine side chains.

The mechanism by which Sec13/31 binding accelerates GTP hydrolysis tenfold⁸ awaits further structural and biochemical analysis. A C-terminal region of the 140K Sec31 molecule, encompassing a proline-rich element, is thought to interact with both Sec23 and Sec24 (ref. 45). A potential binding site on Sec23 for Sec31 is the membrane-distal surface of the gelsolin-like domain (Fig. 3a), on the basis of two features: first, the radial extension of this domain away from the membrane gives it the appearance of a ‘handle’ for Sec31 binding; more directly, its close resemblance to a gelsolin module and the available gelsolin–actin complex structure³⁹ suggests an interaction site centred on helix αQ on the acidic outer surface (Fig. 3a). It will be interesting to discover the precise Sec31 binding site and whether rate acceleration involves direct or allosteric interaction with Sar1–GTP.

This analysis of the Sec23/24–Sar1 structure establishes the molecular basis for GTP-dependent recruitment and formation of a pre-budding complex, and for coat-dependent GTP hydrolysis that controls uncoating. And it provides the foundation for molecular-level studies of the mechanisms of SNARE and cargo selection by the COPII coat complex. □

Methods

Details of protein expression and purification are in Supplementary Information.

Crystallization

The Sec23/24 complex was crystallized at 22 °C by the hanging-drop method, by adding 1 μ l of 35 mg ml^{−1} protein solution containing 15% (w/v) xylitol to 1 μ l of ‘mix’ solution comprising 1.0 M ammonium formate, 18 mM dithiothreitol and 0.1 M HEPES buffer at pH 7.5, and placing this over well solution containing 1.9 M ammonium formate and 0.1 M HEPES at pH 7.5. The crystals are orthorhombic, space group P2₁2₁2₁ ($a = 90.3$ Å, $b = 126.4$ Å, $c = 180.2$ Å). Crystals were transferred to a cryoprotection solution of 6.0 M sodium formate, 10% xylitol and 0.05 M Hepes at pH 7.5, and flash-frozen in liquid propane. For crystallization of the Sec23–Sar1 complex, equimolar Sec23 and Sar1–GppNHP (prepared as described¹⁹) were mixed to give a 44 mg ml^{−1} protein solution. We

mixed 1 μ l with 1 μ l of well solution comprising 12% (w/v) PEG 1500, 10% (v/v) isopropanol, 0.2 M ammonium acetate and 50 mM MES buffer at pH 6.5. Orthorhombic ($P2_12_12_1$) crystals grow with two copies of the Sec23–Sar1 complex in the asymmetric unit (cell dimensions $a = 47.2$ Å, $b = 151.2$ Å, $c = 271.6$ Å). Before flash-freezing, crystals were cryoprotected by transfer to well solution containing additional 15% (v/v) ethylene glycol.

Structure determination

Diffraction data for Sec23/24 were measured at beamline X-25 of the National Synchrotron Light Source (NSLS). Data were processed with programs DENZO and SCALEPACK⁴⁶ (Supplementary Information Table 1). The structure of Sec23/24 was solved by molecular replacement with the program AMORE⁴⁸ using Sec23 as the search model. A translation function (8.0–4.0 Å resolution) gave a single unambiguous (2σ) peak. The model was improved using rigid-body and positional refinement, and a model for Sec24 was built in the resulting electron density map. Further refinement with CNS⁴⁹ yielded a final R -factor of 20.5% ($R_{\text{free}} = 25.3\%$) for data between 25.0 and 2.75 Å resolution. As is found in the Sec23/Sar1 structure, residues Val 30 and Phe 659 are the only Ramachandran plot outliers. The final model is complete with the exception of the disordered Sec23 residues 199–220, 466–480, 543–546 and 728–749, and Sec24 residues 1–132, 363–371, 463–466, 770–778, 823–847 and 876–887.

MAD data for the Sec23–Sar1 complex were collected at beamline F-2 of the Cornell High Energy Synchrotron Source (CHESS) (Supplementary Information Table 1). Data were processed as before, and MAD analysis was done with the program SOLVE⁴⁷ using data between 25.0 and 2.7 Å resolution. SOLVE found 23 of the 32 selenium sites plus two zinc sites, and yielded a mean figure of merit of 0.31 (0.17 for the highest-resolution bin). Density modification was carried out, including non-crystallographic symmetry averaging, with program DM (CCP4 suite⁴⁸). This produced a high-quality electron density map, and a nearly complete model was built with the aid of the known structure of ARF1–GppNHP¹⁹. Refinement of atomic positions and temperature factors, together with a bulk solvent correction, with the program CNS⁴⁹ reduced the R -factor to a final value of 23.8% ($R_{\text{free}} = 29.5\%$) for data between 25.0 and 2.5 Å resolution. The model comprises 13,885 protein atoms, two molecules of Mg^{2+} –GppNHP, two zinc atoms and 171 water molecules. There are four Ramachandran plot outliers: residues Val 30 and Phe 659 in each copy of Sec23. Inspection of electron density maps confirms that these residues are modelled correctly, and adopt high-energy conformations owing to their location in tight turns. The following residues have not been modelled because of weak or no associated electron density: in each molecule of Sec23, residues 96–98, 204–218, 466–482, 732–742; and in each Sar1, residues 157–159.

Received 27 May; accepted 19 July 2002; doi:10.1038/nature01040.

- Mellman, I. & Warren, G. The road taken: past and future foundations of membrane traffic. *Cell* **100**, 99–112 (2000).
- Kirchhausen, T. Three ways to make a vesicle. *Nature Rev. Mol. Cell Biol.* **1**, 187–198 (2000).
- Serafini, T. *et al.* ADP-ribosylation factor is a subunit of the coat of Golgi-derived COP-coated vesicles: a novel role for a GTP-binding protein. *Cell* **67**, 239–253 (1991).
- Donaldson, J. G., Cassel, D., Kahn, R. A. & Klausner, R. D. ADP-ribosylation factor, a small GTP-binding protein, is required for binding of the coatamer protein β -COP to Golgi membranes. *Proc. Natl Acad. Sci. USA* **89**, 6408–6412 (1992).
- Barlowe, C. *et al.* COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907 (1994).
- Tanigawa, G. *et al.* Hydrolysis of bound GTP by ARF protein triggers uncoating of Golgi-derived COP-coated vesicles. *J. Cell Biol.* **123**, 1365–1371 (1993).
- Yoshihisa, T., Barlowe, C. & Schekman, R. Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. *Science* **259**, 1466–1468 (1993).
- Antony, B., Madden, D., Hamamoto, S., Orci, L. & Schekman, R. Dynamics of the COPII coat with GTP and stable analogues. *Nature Cell Biol.* **3**, 531–537 (2001).
- Nagahama, M. *et al.* A v-SNARE implicated in intra-Golgi transport. *J. Cell Biol.* **133**, 507–516 (1996).
- Springer, S. & Schekman, R. Nucleation of COPII vesicular coat complex by endoplasmic reticulum to Golgi vesicle SNAREs. *Science* **281**, 698–700 (1998).
- Grabowski, R. & Gallwitz, D. High-affinity binding of the yeast *cis*-Golgi t-SNARE, Sed5p, to wild-type and mutant Sly1p, a modulator of transport vesicle docking. *FEBS Lett.* **411**, 169–172 (1997).
- Matsuoka, K. *et al.* COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. *Cell* **93**, 263–275 (1998).
- Nakano, A., Brada, D. & Schekman, R. A membrane glycoprotein, Sec12p, required for protein transport from the endoplasmic reticulum to the Golgi apparatus in yeast. *J. Cell Biol.* **107**, 851–863 (1988).
- Barlowe, C. & Schekman, R. SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. *Nature* **365**, 347–349 (1993).
- Weissman, J. T., Plutner, H. & Balch, W. E. The mammalian guanine nucleotide exchange factor mSec12 is essential for activation of the Sar1 GTPase directing endoplasmic reticulum export. *Traffic* **2**, 465–475 (2001).
- Huang, M. *et al.* Crystal structure of Sar1-GDP at 1.7 Å resolution and the role of the NH2 terminus in ER export. *J. Cell Biol.* **155**, 937–948 (2001).
- Amor, J. C., Harrison, D. H., Kahn, R. A. & Ringe, D. Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* **372**, 704–708 (1994).
- Antony, B., Beraud-Dufour, S., Chardin, P. & Chabre, M. N-terminal hydrophobic residues of the G-protein ADP-ribosylation factor-1 insert into membrane phospholipids upon GDP to GTP exchange. *Biochemistry* **36**, 4675–4684 (1997).
- Goldberg, J. Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell* **95**, 237–248 (1998).
- Beraud-Dufour, S., Paris, S., Chabre, M. & Antony, B. Dual interaction of ADP ribosylation factor 1 with Sec7 domain and with lipid membranes during catalysis of guanine nucleotide exchange. *J. Biol. Chem.* **274**, 37629–37636 (1999).
- Lederkremer, G. Z. *et al.* Structure of the Sec23p/24p and Sec13p/31p complexes of COPII. *Proc. Natl Acad. Sci. USA* **98**, 10704–10709 (2001).
- Matsuoka, K., Schekman, R., Orci, L. & Heuser, J. E. Surface structure of the COPII-coated vesicle. *Proc. Natl Acad. Sci. USA* **98**, 13705–13709 (2001).
- Kuehn, M. J., Herrmann, J. M. & Schekman, R. COPII-cargo interactions direct protein sorting into ER-derived transport vesicles. *Nature* **391**, 187–190 (1998).
- Aridor, M., Weissman, J., Bannykh, S., Nuoffer, C. & Balch, W. E. Cargo selection by the COPII budding machinery during export from the ER. *J. Cell Biol.* **141**, 61–70 (1998).
- Peng, R., Grabowski, R., De Antoni, A. & Gallwitz, D. Specific interaction of the yeast *cis*-Golgi syntaxin Sed5p and the coat protein complex II component Sec24p of endoplasmic reticulum-derived transport vesicles. *Proc. Natl Acad. Sci. USA* **96**, 3751–3756 (1999).
- Votsmeier, C. & Gallwitz, D. An acidic sequence of a putative yeast Golgi membrane protein binds COPII and facilitates ER export. *EMBO J.* **20**, 6742–6750 (2001).
- Martinez-Menarguez, J. A., Geuze, H. J., Slot, J. W. & Klumperman, J. Vesicular tubular clusters between the ER and Golgi mediate concentration of soluble secretory proteins by exclusion from COPI-coated vesicles. *Cell* **98**, 81–90 (1999).
- Balch, W. E., McCaffery, J. M., Plutner, H. & Farquhar, M. G. Vesicular stomatitis virus glycoprotein is sorted and concentrated during export from the endoplasmic reticulum. *Cell* **76**, 841–852 (1994).
- ter Haar, E., Musacchio, A., Harrison, S. C. & Kirchhausen, T. Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. *Cell* **95**, 563–573 (1998).
- Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. Molecular architecture and functional model of the endocytic AP2 complex. *Cell* **109**, 523–535 (2002).
- Huang, M., Weissman, J. T., Wang, C., Balch, W. E. & Wilson, I. A. Protein engineering for crystallization of the GTPase Sar1 that regulates ER vesicle budding. *Acta Crystallogr. D* **58**, 700–703 (2002).
- Paris, S. *et al.* Role of protein–phospholipid interactions in the activation of ARF1 by the guanine nucleotide exchange factor Arno. *J. Biol. Chem.* **272**, 22221–22226 (1997).
- Goldberg, J. Structural and functional analysis of the ARF1–ARFGAP complex reveals a role for coatamer in GTP hydrolysis. *Cell* **96**, 893–902 (1999).
- Pasqualato, S., Menetrey, J., Franco, M. & Cherfils, J. The structural GDP/GTP cycle of human Arf6. *EMBO Rep.* **2**, 234–238 (2001).
- Hanzal-Bayer, M., Renault, L., Roversi, P., Wittinghofer, A. & Hillig, R. C. The complex of Arl2-GTP and PDE δ : from structure to function. *EMBO J.* **21**, 2095–2106 (2002).
- Shimoni, Y. *et al.* Lst1p and Sec24p cooperate in sorting of the plasma membrane ATPase into COPII vesicles in *Saccharomyces cerevisiae*. *J. Cell Biol.* **151**, 973–984 (2000).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745 (1999).
- Robinson, R. C. *et al.* Domain movement in gelsolin: a calcium-activated switch. *Science* **286**, 1939–1942 (1999).
- Peng, R., De Antoni, A. & Gallwitz, D. Evidence for overlapping and distinct functions in protein transport of coat protein Sec24p family members. *J. Biol. Chem.* **275**, 11521–11528 (2000).
- Franco, M., Chardin, P., Chabre, M. & Paris, S. Myristoylation is not required for GTP-dependent binding of ADP-ribosylation factor ARF1 to phospholipids. *J. Biol. Chem.* **268**, 24531–24534 (1993).
- Franco, M., Chardin, P., Chabre, M. & Paris, S. Myristoylation-facilitated binding of the G protein ARF1GDP to membrane phospholipids is required for its activation by a soluble nucleotide exchange factor. *J. Biol. Chem.* **271**, 1573–1578 (1996).
- Randazzo, P. A. Functional interaction of ADP-ribosylation factor 1 with phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* **272**, 7688–7692 (1997).
- Scheffzek, K., Ahmadian, M. R. & Wittinghofer, A. GTPase-activating proteins: helping hands to complement an active site. *Trends Biochem. Sci.* **23**, 257–262 (1998).
- Shaywitz, D. A., Espenshade, P. J., Gimeno, R. E. & Kaiser, C. A. COPII subunit interactions in the assembly of the vesicle coat. *J. Biol. Chem.* **272**, 25413–25416 (1997).
- Otwinoski, W. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Tervilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Collaborative Computational Project Number 4 The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank E. Mossessova for the Sar1 expression construct, J. Walker for assistance and advice on data collection and processing, C. Heaton for use of synchrotron facilities at CHESS, and M. Becker at NSLS. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, and the Pew Scholars Program in the Biomedical Sciences.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.G. (e-mail: jonathan@ximpact4.ski.mskcc.org). Coordinates have been deposited with the Protein Data Bank under accession codes 1M20 and 1M2V.

Allowed and forbidden transitions in artificial hydrogen and helium atoms

Toshimasa Fujisawa*, David Guy Austing*†, Yasuhiro Tokura*, Yoshiro Hirayama*‡ & Seigo Tarucha*§||

* NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, 243-0198, Japan

† Institute for Microstructural Sciences M23A, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

‡ CREST Interacting Carrier Electron Project, 4-1-8 Honmachi, Kawaguchi, 331-0012, Japan

§ University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan

|| ERATO Mesoscopic Correlation Project, 3-1 Morinosato-Wakamiya, Atsugi, 243-0198, Japan

The strength of radiative transitions in atoms is governed by selection rules that depend on the occupation of atomic orbitals with electrons¹. Experiments have shown^{2–5} similar electron occupation of the quantized energy levels in semiconductor quantum dots—often described as artificial atoms. But unlike real atoms, the confinement potential of quantum dots is anisotropic, and the electrons can easily couple with phonons of the material⁶. Here we report electrical pump-and-probe experiments that probe the allowed and ‘forbidden’ transitions between energy levels under phonon emission in quantum dots with one or two electrons (artificial hydrogen and helium atoms). The forbidden transitions are in fact allowed by higher-order processes where electrons flip their spin. We find that the relaxation time is about 200 μ s for forbidden transitions, 4 to 5 orders of magnitude longer than for allowed transitions. This indicates that the spin degree of freedom is well separated from the orbital degree of freedom, and that the total spin in the quantum dots is an excellent quantum number. This is an encouraging result for potential applications of quantum dots as basic entities for spin-based quantum information storage.

The quantum dot (QD) that we study is located in a circular pillar (diameter 0.5 μ m) fabricated from an AlGaAs/InGaAs heterostructure (Fig. 1a, b and ref. 4). Electrons are confined in an In_{0.05}Ga_{0.95}As quantum well (thickness $a = 12$ nm) in the vertical (z) direction, and approximately by two orthogonal two-dimensional harmonic potentials in the lateral (x and y) direction (corresponding confinement energies, $\hbar\omega_x \approx 2.5$ meV and $\hbar\omega_y \approx 5.5$ meV, where $\hbar$ is Planck’s constant divided by 2π)^{7,8}. As our QD does not have circular symmetry, orbital degeneracy is lifted even at zero magnetic field, and only a twofold spin degeneracy is expected. This non-circularity does not much affect our discussion, and we still use 1s and 2p to label the orbitals for convenience (Fig. 1c).

First we investigate the $N = 1$ QD (artificial hydrogen), in which a single electron occupies the 1s orbital (the ground state) or the 2p orbital (the first excited state). The energy spectrum of these states can be obtained by tunnelling spectroscopy, in which a peak in the derivative of the (source–drain) current with respect to the gate voltage, dI_{sd}/dV_g , appears each time an empty dot state enters the transport window⁵. A colour plot of dI_{sd}/dV_g versus V_g traces taken as a function of magnetic field (B) applied in the z direction is shown in Fig. 1c. The peak spacing within a stripe can be related to the energy spacings between corresponding states. The energy spacing between the 1s and 2p states, ϵ_{1s-2p} , of the $N = 1$ QD deduced from the first current stripe is plotted in Fig. 2e.

We now focus on the energy relaxation from the 2p state to the 1s state in the $N = 1$ QD, which changes the electron’s orbital momentum but preserves the spin. Electrical pump-and-probe experiments are performed by applying a time-dependent gate voltage, $V_g(t)$, which switches between V_l and V_h (Fig. 2a).

Experimental details are given in refs 9 and 10. First, the $N = 0$ QD is prepared during the low-phase of the pulse ($V_g = V_l$; Fig. 2b). The period $t_l = 100$ –200 ns, is made long enough to ensure that both the 1s and 2p states are empty. When the pulse is switched on ($V_g = V_h$; Fig. 2c), such that only the 2p state is located in the transport window, an electron can be injected into that state from the source (pump) with a time constant $\Gamma_s^{-1} \approx 3$ ns. The electron can only escape to the drain (probe) more slowly, with a time constant $\Gamma_d^{-1} \approx 100$ ns. However, this escape process can be interrupted by the relaxation into the 1s ground state. Thus, the current contains information about the relaxation lifetime, τ_{1s-2p} . We measure the averaged direct current, I_p , under the application of the pulse train. Figure 2d shows how this current changes with the pulse length, t_h . I_p is then converted into an average number of tunnelling electrons per pulse, $\langle n_t \rangle = I_p(t_h + t_l)/e$ (e is the elementary charge). From a detailed analysis of the rate equations including all possible tunnelling processes, we find $\langle n_t \rangle \approx \Gamma_d \tau_{1s-2p} [1 - \exp(-t_h/\tau_{1s-2p})]$ under the condition $\Gamma_s^{-1} \lesssim \tau_{1s-2p} < \Gamma_d^{-1}$, required for the relaxation time measurement⁹. We made sure that this condition was satisfied in all measurements for the $N = 1$ QD. The relaxation time thus estimated from the rise time of $\langle n_t \rangle$ is $\tau_{1s-2p} \approx 10$ ns for the case in Fig. 2d. The small saturation value of $\langle n_t \rangle$ (~ 0.02) indicates very efficient relaxation.

Our observations for the $N = 1$ QD are consistent with momentum relaxation dominated by spontaneous emission of acoustic phonons at low temperature⁶. Because of the discrete energy of the states, relaxation involves emission of a phonon of energy ϵ_{1s-2p} (with corresponding wavelength λ_{1s-2p}). In our experiment, ϵ_{1s-2p} , and hence λ_{1s-2p} , vary with B (Fig. 2e, f). Note also that the characteristic sizes l_x and l_y for the lateral dimensions of the QD decrease with increasing B . The strength of the electron–phonon interaction is expected to be suppressed for values of λ_{1s-2p} that are

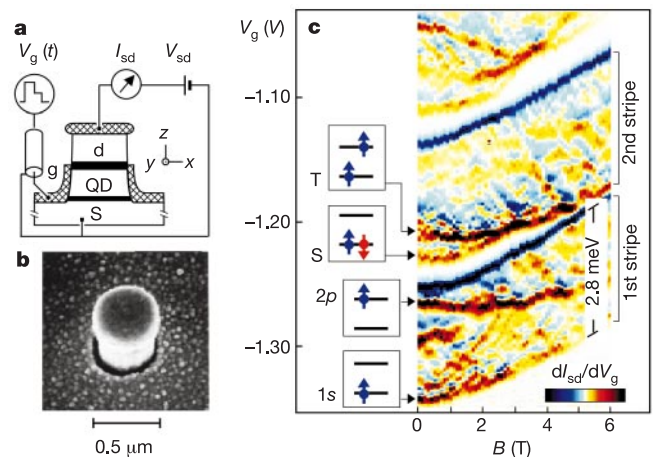


Figure 1 Artificial hydrogen and helium atoms. **a**, Schematic set-up for pulse measurements on the vertical quantum dot (QD). The In_{0.05}Ga_{0.95}As QD is connected to source (s) and drain (d) electrodes made of Si-doped GaAs by asymmetric Al_{0.22}Ga_{0.78}As tunnelling barriers (lower barrier, 7 nm thick; upper barrier, 8.5 nm thick). The tunnelling rates through the barriers, $\Gamma_s \approx (3 \text{ ns})^{-1}$ and $\Gamma_d \approx (100 \text{ ns})^{-1}$, are obtained by separate measurements. The surrounding gate electrode (g) is connected to a pulse generator, which produces a gate voltage, $V_g(t)$, of rectangular or double-step shape. The measurements are performed in a dilution refrigerator at a temperature, T , of ~ 100 mK, unless otherwise stated, in a magnetic field $B = 0$ –5 T applied parallel to the z direction. **b**, Scanning electron micrograph of a control device. **c**, dI_{sd}/dV_g for the $N = 1$ and 2 QD taken with a large source–drain voltage ($V_{sd} = 2.8$ mV). The first (second) stripe gives information about the $N = 1$ (2) QD. The peaks indicated by the arrows show the B -field evolution of the ground state (lowest edge of each stripe) and the first excited state. The relevant 1- and 2- electron configurations are also shown, in which the lower (upper) horizontal line represents the 1s (2p) orbital.

smaller than the characteristic size of the QD (phonon bottleneck effect¹¹). Therefore, the B dependence of τ_{1s-2p} in Fig. 2g is because the phonon emission is suppressed (that is, τ_{1s-2p} increases) with decreasing B when λ_{1s-2p} becomes shorter than a , l_x and l_y .

In order to be quantitative, we calculate the phonon emission rate from Fermi's golden rule including both deformation and piezo-electric coupling with standard GaAs material parameters^{12,13}. For simplicity, the calculation is done for a circular dot, whose effective confinement energy is $\hbar\omega_{\text{eff}} = \hbar\sqrt{\omega_x\omega_y(1 + \omega_c^2/(\omega_x + \omega_y)^2)}$, where ω_c is the cyclotron frequency⁷. As shown by the solid line in Fig. 2g, we find agreement with the data. The difference (by about a factor of 2 or 3) might come from the assumptions about the confinement potential and uncertainties in the material parameters. Thus, the fast energy relaxation in the $N = 1$ QD can be well understood by spontaneous emission of a phonon.

In contrast, the relaxation time is very different for the $N = 2$ QD (artificial helium). At low magnetic fields (see second stripe in Fig. 1c for $B < 2.5$ T), the many-body ground state is a spin singlet (labelled S) with two antiparallel-spin electrons occupying the 1s

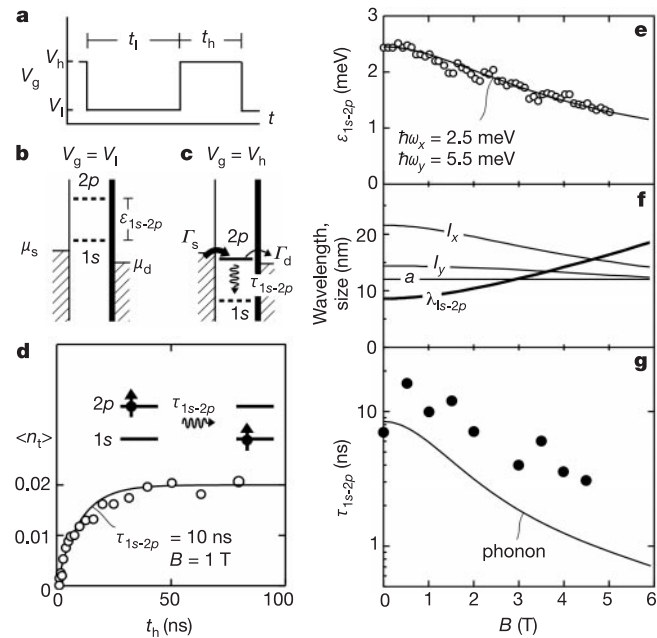


Figure 2 Relaxation time of a one-electron QD (artificial hydrogen atom). **a**, Pulse waveform used for the electrical pump-and-probe experiment (l, low; h, high). **b** and **c**, Schematic energy diagrams along the z direction showing low and high pulse situations. The thick and thin vertical lines denote the asymmetric tunnelling barriers. States in the electrodes are filled up to the Fermi energies, μ_s for the source and μ_d for the drain. The source-drain voltage, V_{sd} , opens a small transport window $eV_{sd} = \mu_s - \mu_d \approx 0.1$ meV. Solid and dashed horizontal lines denote filled and empty single-particle states, respectively. When $V_g = V_l$ (**b**), the 1s and 2p states are located above μ_s and μ_d . When $V_g = V_h$ (**c**), only the 2p state is located in the transport window. The 2p state is pumped from the source at a tunnelling rate, $\Gamma_s \approx (3 \text{ ns})^{-1}$, and probed at a slower rate, $\Gamma_d \approx (100 \text{ ns})^{-1}$. The current measures the momentum relaxation time of the 2p state, τ_{1s-2p} . **d**, The average number of tunnelling electrons per pulse, $\langle n_t \rangle$, measured at 1 T. The relaxation time, $\tau_{1s-2p} = 10$ ns, is obtained from the exponential curve (solid line) fitted to the data. The inset shows the electron configuration before and after relaxation. **e–g**, Magnetic field (B) dependence of **e**, the energy spacing between the 2p excited state and the 1s ground state, ϵ_{1s-2p} , **f**, the longitudinal acoustic photon wavelength, λ_{1s-2p} , and characteristic sizes of the QD (a , l_x and l_y), and **g**, the energy relaxation time, τ_{1s-2p} . The solid line in **e** is a fitted curve with elliptic confinement energies, $\hbar\omega_x \approx 2.5$ meV and $\hbar\omega_y \approx 5.5$ meV. In **f**, λ_{1s-2p} is calculated for the phonon at energy ϵ_{1s-2p} using a GaAs sound velocity of $5,100 \text{ m s}^{-1}$. The characteristic lateral size in the x/y direction is given by $l_{x/y} = \sqrt{\hbar/m^*(\omega_{x/y}^2 + \omega_c^2/4)^{-1/4}}$, where m^* is the effective mass. The solid line in **g** is calculated for spontaneous emission of an acoustic phonon.

orbital, while the first excited state is a spin triplet (labelled T) with two parallel-spin electrons, one each occupying the 1s and 2p orbitals^{4,5}. Because of Coulomb interactions, the energy spacing between the two states, ϵ_{S-T} (~ 0.6 meV at $B = 0$ T), is smaller than ϵ_{1s-2p} . Energy relaxation from the first excited state (T) to the ground state (S) not only involves the same change in orbital momentum as that in the $N = 1$ QD, but also requires a spin flip because of Pauli exclusion. A simple phonon-emission transition from the triplet to the singlet is forbidden by spin conservation.

We now investigate to what degree this transition is 'forbidden'. The simple rectangular pulse technique used for the $N = 1$ QD is not useful for this $N = 2$ QD transition, because the relaxation lifetime, τ_{S-T} , is always beyond the measurable range ($\tau_{S-T} > \Gamma_d^{-1} \approx 100 \text{ ns}$). Instead, we subject the QD to a double-step voltage pulse, in which V_g is switched between three voltages, V_l , V_h and V_m (Fig. 3a). First, when $V_g = V_l$ (Fig. 3b), the $N = 1$ QD is prepared during a sufficiently long period, $t_l = 100$ ns. When V_g is suddenly increased to V_h (Fig. 3c), an electron can enter to create the $N = 2$ triplet state within the interval $\sim \Gamma_s^{-1} = 3\text{--}7$ ns. $V_g = V_h$ for the duration $t_h = 100 \text{ ns--}100 \mu\text{s}$, which is much longer than Γ_s^{-1} . The triplet may experience a relaxation process during this time. When V_g is changed to V_m (Fig. 3d), an electron in the triplet state can tunnel out to the drain, if the triplet state has not yet

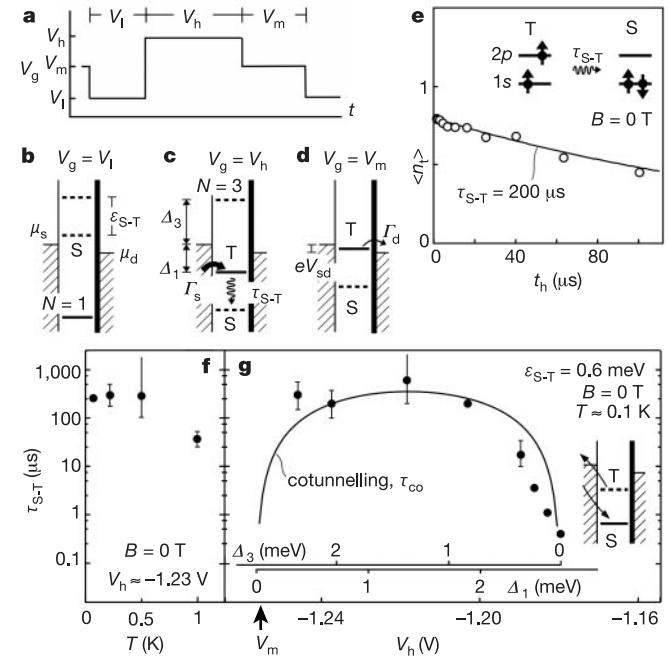


Figure 3 Relaxation time of a two-electron QD (artificial helium atom). **a**, Double-step pulse waveform to measure extremely long relaxation times. **b–d**, Schematic energy diagrams showing low, high and intermediate pulse situations. Solid and dashed horizontal lines denote filled and empty many-body states, respectively. $\mu_s - \mu_d \approx 0.1$ meV. When $V_g = V_l$ (**b**), the spin-singlet ground state S and the spin-triplet first excited state T are located above μ_s and μ_d . The system will always become the $N = 1$ QD after a sufficiently long period, $t_l = 100$ ns. When $V_g = V_h$ (**c**), the QD can be excited to the triplet state within $\Gamma_s^{-1} \approx 7$ ns. The triplet state can then relax to the singlet state during the period, $t_h = 0.1\text{--}100 \mu\text{s}$. When $V_g = V_m$ (**d**), the triplet state is probed by allowing an electron to tunnel into the drain. This period is fixed at $t_m = 300$ ns. **e**, Average number of tunnelling electrons per pulse, $\langle n_t \rangle$, at 0 T. The relaxation time, $\tau_{S-T} = 200 \mu\text{s}$, is obtained from the exponential decay (solid line). Inset, electron configuration before and after relaxation. **f**, Temperature (T) dependence of the relaxation time τ_{S-T} at 0 T. **g**, The gate voltage (V_h) dependence of τ_{S-T} . V_h is also converted into Δ_1 and Δ_3 energy scales. Δ_1 and Δ_3 are indicated in **c**. The solid line is calculated for cotunnelling processes. Inset, diagram of these inelastic cotunnelling processes.

relaxed to the singlet state. We set the period, $t_m = 300$ ns, to be longer than Γ_d^{-1} so we can read out the signal. We repeatedly apply the double-step pulse (effectively $\sim 10^7$ times) to obtain a reliable current I_p , and evaluate the average number of tunnelling electrons $\langle n_t \rangle = I_p(t_l + t_h + t_m)/e$. As the current measures the unrelaxed electron number, $\langle n_t \rangle = A \exp(-t_h/\tau_{S-T})$ for the condition $\Gamma_s^{-1} \leq \tau_{S-T}$ (no upper limit in principle), where $A \approx 1$ is approximately the ratio of Γ_s^{-1} for the triplet to that for the singlet. Figure 3e shows a typical measurement of $\langle n_t \rangle$ at 0 T, indicating a relaxation time of $\tau_{S-T} \approx 200$ μ s. This relaxation time is 4 to 5 orders of magnitude longer than that observed in the $N = 1$ QD.

Looking at the properties of τ_{S-T} , we find no clear B dependence (always longer than 100 μ s), at least for ε_{S-T} between 0.6 meV at $B = 0$ T and 0.24 meV at $B = 2$ T (not shown). We also investigate the temperature dependence (Fig. 3f), but not clear change is observed up to 0.5 K. τ_{S-T} decreases above 0.5 K, where thermal excitation from the QD to the electrodes becomes important. On the other hand, we do find that τ_{S-T} strongly depends on the high gate-pulse voltage V_h , during which relaxation takes place (Fig. 3g). Although V_h is deep in the $N = 2$ Coulomb blockade region (-1.27 V $< V_h < -1.16$ V), τ_{S-T} decreases rapidly at $V_h \approx -1.18$ V. This V_h dependence implies a strong influence of the electrodes. Even though the Coulomb blockade is robust in the suppression of transport, higher-order tunnelling processes can contribute to the relaxation. An electron in the dot can be replaced with an electron of opposite spin from the electrodes (see Fig. 3g inset). This results in energy relaxation in the QD, and the electrode gains the same energy. This inelastic cotunnelling rate, τ_{co}^{-1} , is estimated by considering second-order tunnelling processes^{14–16}. For the relaxation mechanisms considered here, the $N = 2$ QD can relax virtually through $N = 1$ or $N = 3$ intermediate states. Note that this process does not cause a net current, even at a finite voltage of $e|V_{sd}| < \varepsilon_{S-T}$ (ref. 17). Assuming $V_{sd} = 0$ V and zero temperature for simplicity, we obtain $\tau_{co}^{-1} = \varepsilon_{S-T}(\hbar\Gamma_s + \hbar\Gamma_d)^2(\Delta_1^{-1} + \Delta_3^{-1})^2/h$. Here, Δ_1 and Δ_3 are respectively the energies required to excite the initial $N = 2$ triplet state to the $N = 1$ and 3 intermediate states (Fig. 3c). We can extract Δ_1 and Δ_3 from V_h , and the values are shown in Fig. 3g. The solid line shows τ_{co} , the relaxation time due to cotunnelling, calculated with experimentally deduced parameters ($\varepsilon_{S-T} = 0.6$ meV, and $(\Gamma_s + \Gamma_d)^{-1} = 7$ ns). The observed relaxation time can be well understood by inelastic cotunnelling.

Our observations for $N = 1$ and 2 QDs can be compared with real atoms¹. The transition from the $2p$ state to the $1s$ state in atomic hydrogen is allowed by photon emission, whereas that in artificial hydrogen is allowed by phonon emission. The transition from the spin-triplet state to the spin-singlet state is forbidden by conservation of the total spin for both atomic helium and artificial helium. The difference between the allowed and forbidden transitions leads to more than 11 orders of magnitude difference in the relaxation times for real hydrogen and helium atoms. Our observation of 4 to 5 orders of magnitude difference in artificial atoms is not as high, but is still surprisingly large. (Note that this difference would become larger if the cotunnelling could be suppressed by using thicker tunnelling barriers.) Very importantly, the large difference between τ_{1s-2p} and τ_{S-T} originates from the fact that other effects, such as spin-orbit and hyperfine interactions^{18,19}, must have only a weak effect on the breaking of the ‘forbidden’ symmetries. We now discuss how small these hidden contributions are by focusing on spin-orbit interactions.

Spin-orbit interactions are predicted to give the dominant contribution to spin relaxation in GaAs QD systems¹⁹, although this is still an extremely small effect. For simplicity, we consider the spin-orbit interaction energy, Δ_{so} , only for coupling between the $1s$ and $2p$ orbitals, but include all effects that mix spin and orbital degrees of freedom. Simple perturbation theory²⁰ predicts that the relaxation time from the triplet to the singlet is given by $\tau_{S-T,so} \approx (\varepsilon_{S-T}/\Delta_{so})^2\tau_{\text{phonon}}(\varepsilon_{S-T})$. Here, $\tau_{\text{phonon}}^{-1}(\varepsilon_{S-T})$ is the phonon emis-

sion rate at the phonon energy, ε_{S-T} , and we know that τ_{phonon} is well accounted for by the electron-phonon interaction. Therefore, we can deduce an upper bound of $\Delta_{so} < 4$ μ eV from our observations ($\tau_{S-T} > 200$ μ s). This value is close to the spin splitting energy (~ 2.5 μ eV) observed in a GaAs two-dimensional electron gas system²¹.

Our experiments indicate that the spin degree of freedom in QDs is well separated from the orbital degree of freedom. This is particularly attractive for applications to spin memories and spin quantum bits (qubits)^{22–24}. For a simple scheme involving just a single-electron spin in a magnetic field, the spin-orbit interactions can affect the energy relaxation time (T_1) of a spin qubit. We estimate the dominant contribution, $T_{1,so}$, using a perturbative approach: $T_{1,so} \approx (\varepsilon_{1s-2p}/\Delta_{so})^2\tau_{\text{phonon}}(\varepsilon_Z)$. As $\Delta_{so} < 4$ μ eV, this yields $T_{1,so} > 1$ ms for a Zeeman splitting (unresolved in our measurement) $\varepsilon_Z \approx 0.1$ meV and $\varepsilon_{1s-2p} \approx 1.2$ meV at $B \approx 5$ T ($T_{1,so} > 100$ μ s at $B \approx 9$ T). This $T_{1,so}$ is thus comparable to that obtained by electron spin resonance for donor states in GaAs (ref. 25), and is much longer than the time required for typical one- and two-qubit operations²⁶. Note that small spin-orbit interactions are also desirable with respect to the dephasing time (T_2) of a spin qubit²⁷. Our results therefore encourage further research on the use of the spin degree of freedom in QDs. □

Received 2 May; accepted 5 July 2002; doi:10.1038/nature00976.

1. Bethe, H. A. & Salpeter, E. E. *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, Berlin, 1957).
2. Kouwenhoven, L. P., et al. in *Mesoscopic Electron Transport* NATO ASI series E 345 (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) 105–214 (Kluwer, Dordrecht, 1997).
3. Ashoori, R. C. et al. Single-electron capacitance spectroscopy of discrete quantum levels. *Phys. Rev. Lett.* **68**, 3088–3091 (1992).
4. Tarucha, S., Austing, D. G., Honda, T., van der Hage, R. J. & Kouwenhoven, L. P. Shell filling and spin effects in a few electron quantum dot. *Phys. Rev. Lett.* **77**, 3613–3616 (1996).
5. Kouwenhoven, L. P. et al. Excitation spectra of circular, few-electron quantum dots. *Science* **278**, 1788–1792 (1997).
6. Fujisawa, T. et al. Spontaneous emission spectrum in double quantum dot devices. *Science* **282**, 932–935 (1998).
7. Tokura, Y., Sasaki, S., Austing, D. G. & Tarucha, S. Excitation spectra and exchange interactions in circular and elliptical quantum dots. *Physica B* **298**, 260–264 (2001).
8. Matagne, P., Leburton, J. P., Austing, D. G. & Tarucha, S. Shell charging and spin-filling sequences in realistic vertical quantum dots. *Phys. Rev. B* **65**, 085325 (2002).
9. Fujisawa, T., Tokura, Y. & Hirayama, Y. Transient current spectroscopy of a quantum dot in the Coulomb blockade regime. *Phys. Rev. B* **63**, 081304 (2001).
10. Fujisawa, T., Austing, D. G., Tokura, Y., Hirayama, Y. & Tarucha, S. Non-equilibrium transport through a vertical quantum dot in the absence of spin-flip energy relaxation. *Phys. Rev. Lett.* **88**, 236802 (2002).
11. Benisty, H., Sotomayer-Torres, C. M. & Weisbuch, C. Intrinsic mechanism for the poor luminescence properties of quantum-box systems. *Phys. Rev. B* **44**, 10945–10948 (1991).
12. Seeger, K. *Semiconductor Physics: An Introduction* 153–213 (Springer, Berlin, 1985).
13. Bockelmann, U. Phonon scattering between zero-dimensional electronic states: Spatial versus Landau quantization. *Phys. Rev. B* **50**, 17271–17279 (1994).
14. Averin, D. V. & Nazarov, Yu. V. in *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* (eds Grabert, H. & Devoret, M. H.) 217–247 (Plenum and NATO Scientific Affairs Division, New York, 1992).
15. Eto, M. Electronic states and transport phenomena in quantum dot systems. *Jpn J. Appl. Phys.* **40**, 1929–1935 (2001).
16. Sukhorukov, E. V., Burkard, G., Loss, D. *Phys. Rev. B* **63**, 125315 (2001).
17. De Franceschi, S. et al. Electron cotunnelling in a semiconductor quantum dot. *Phys. Rev. Lett.* **86**, 878–881 (2001).
18. Pikus, G. E. & Titkov, A. N. in *Optical Orientation* (eds Meier, F. & Zakharchenya, B. P.) 73–131 (Elsevier, Amsterdam, 1984).
19. Khaetskii, A. V. & Nazarov, Yu. V. Spin relaxation in semiconductor quantum dots. *Phys. Rev. B* **61**, 12639–12642 (2000).
20. Halperin, W. P. Quantum size effects in metal particles. *Rev. Mod. Phys.* **58**, 533–606 (1986).
21. Bychkov, Yu. A. & Rashba, E. I. Properties of a 2D electron gas with lifted spectral degeneracy. *JETP Lett.* **39**, 78–81 (1984).
22. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
23. Recher, P., Sukhorukov, E. V. & Loss, D. Quantum dot as spin filter and spin memory. *Phys. Rev. Lett.* **85**, 1962–1965 (2000).
24. Ciorga, M. et al. Readout of a single electron spin based quantum bit by current detection. *Physica E* **11**, 35–40 (2001).
25. Seck, M., Potemski, M. & Wyder, P. High-field spin resonance of weakly bound electrons in GaAs. *Phys. Rev. B* **56**, 7422–7427 (1997).
26. Gupta, J. A., Knobel, R., Samarth, N. & Awschalom, D. D. Ultrafast manipulation of electron spin coherence. *Science* **292**, 2458–2461 (2001).
27. Kavokin, K. V. Anisotropic exchange interaction of localized conduction-band electrons in semiconductors. *Phys. Rev. B* **64**, 075305 (2001).

Acknowledgements

We thank G. E. W. Bauer, T. Honda, T. Inoshita, A. V. Khaetskii and L. P. Kouwenhoven for discussions and help.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.F.

(e-mail: fujisawa@will.brl.ntt.co.jp).

Forward scattering due to slow-down of the intermediate in the $\text{H} + \text{HD} \rightarrow \text{D} + \text{H}_2$ reaction

Steven A. Harich*, Dongxu Dai*†, Chia C. Wang*‡, Xueming Yang*†§, Sheng Der Chao|| & Rex T. Skodje||¶

* Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan

† Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

‡ Department of Chemistry, National Taiwan University, Taipei, Taiwan

§ Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan

|| Institute of Molecular Science, Myodaiji, Okazaki 444-8585, Japan

¶ Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

Quantum dynamical processes near the energy barrier that separates reactants from products influence the detailed mechanism by which elementary chemical reactions occur. In fact, these processes can change the product scattering behaviour from that expected from simple collision considerations, as seen in the two classical reactions $\text{F} + \text{H}_2 \rightarrow \text{HF} + \text{H}$ and $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ and their isotopic variants. In the case of the $\text{F} + \text{HD}$ reaction, the role of a quantized trapped Feshbach resonance state had been directly determined¹, confirming previous conclusions² that Feshbach resonances cause state-specific forward scattering of product molecules. Forward scattering has also been observed in the $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$ reaction^{3,4} and attributed to a time-delayed mechanism^{3,5–7}. But despite extensive experimental^{8–12} and theoretical^{13–18} investigations, the details of the mechanism remain unclear. Here we present crossed-beam scattering experiments and quantum calculations on the $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction. We find that the motion of the system along the reaction coordinate slows down as it approaches the top of the reaction barrier, thereby allowing vibrations perpendicular to the reaction coordinate and forward scattering. The reaction thus proceeds, as previously suggested⁷, through a well-defined ‘quantized bottleneck state’ different from the trapped Feshbach resonance states observed before.

In this work, we studied the $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction experimentally using the $\text{H}(\text{D})$ -atom Rydberg time-of-flight (HRTOF) technique, developed in refs 12 and 19. We measured time-of-flight (TOF) spectra of the D-atom products at different laboratory scattering angles at the collisional energy of 1.200 eV. The TOF spectra measured are then converted to the centre-of-mass product translational energy distributions. Different H_2 -product rovibrational states can be resolved clearly in this experiment. From these distributions, relative quantum state-specific differential cross-sections (DCS) are obtained. On the basis of this measured data, a full three-dimensional (3D) product contour plot is constructed and shown in Fig. 1. One of the most interesting observations is the forward scattering peak seen in Fig. 1, similar to the forward

scattering observed in the $\text{H} + \text{D}_2$ reaction^{3–5} at higher collisional energies. The characteristics of the forward scattering H_2 products are unique. First, it is clear that the forward scattering H_2 -products are significantly colder rotationally, with mainly $\text{H}_2(j' = 1)$ populated, than the backwards and sideways scattering products. Second, the angle range of the observed forward scattering peak is extremely narrow. Experimental characteristics of the forward scattering peak including the rotational distribution and the angular width can be readily reproduced by the quantum calculations. However, to assign the forward peak to specific dynamical behaviour requires further theoretical analysis.

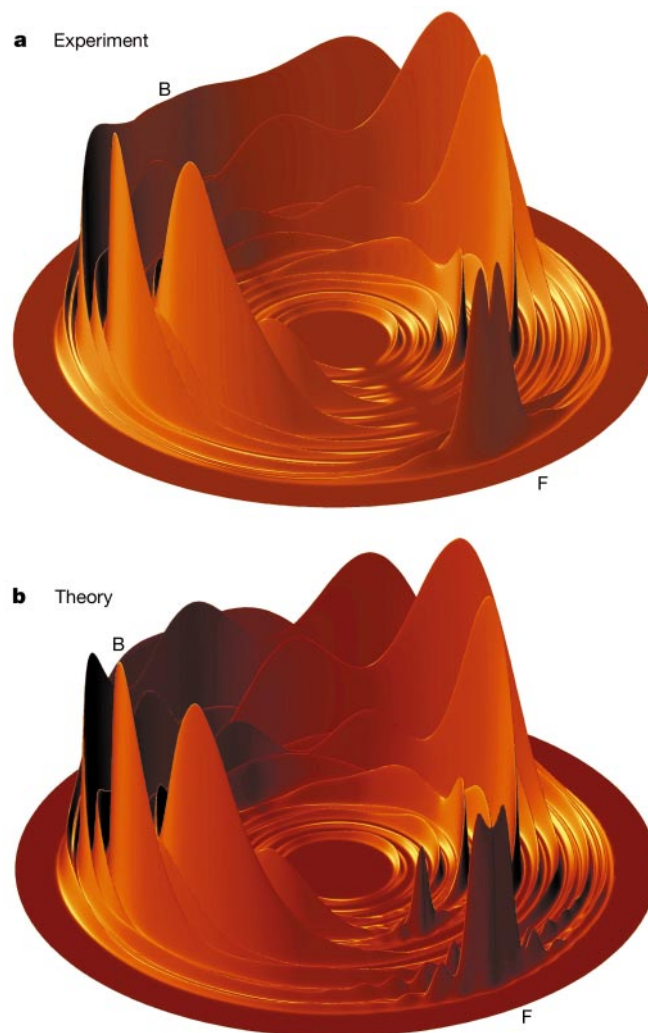


Figure 1 Three-dimensional product contour plots as a function of product velocity. Shown are product contour plots versus product velocity in the centre-of-mass frame obtained experimentally (a) and theoretically (b). Experimentally, time-of-flight spectra of the D-atom product were measured at every 5 degrees and then converted into the translational energy distribution in the centre-of-mass frame. From these translational energy distributions, the angular distributions of all H_2 -product quantum states can be determined. The experimental 3D product contour plot (a) is obtained from fitting smoothly the experimental angular distributions of all product (H_2) quantum states using a multi-order polynomial function. Theoretically, the S -matrix for $\text{H} + \text{HD}$ was calculated using quantum-reactive scattering methods described previously^{17,18}, on the BKMP2 PES²⁰. Differential cross-sections (DCS) at different centre-of-mass angles were computed from the S -matrix. The theoretical 3D product contour plot is constructed from these calculated DCS. The forward-scattering direction ($\theta = 0^\circ$) is denoted F and the backward-scattering direction is denoted B. The height of the peaks represents the magnitude of DCS, while the radius from the centre represents the product translational energy, E_T .

To understand the detailed dynamics observed for this reaction, especially the forward-scattering, we carried out quantum-reactive-scattering calculations^{17,18} on the basis of the BKMP2 PES²⁰ (the improved H₃ potential energy surface by Boothroyd, Keogh, Martin and Peterson). State-specific integral cross-sections (ICS) and DCS were obtained from these calculations. Using the calculated state-specific DCS, a theoretical full 3D product contour plot, also shown in the lower panel in Fig. 1, is constructed in comparison with the experimental result at the collisional energy $E_c = 1.200$ eV. The theoretical results are in good overall agreement with the experimental results, except at very low translational energy in the forward direction, corresponding to the H₂($v' = 1$) products. After careful examinations, the lack of experimental observation for the forward H₂($v' = 1$) products is attributed to a large background in the TOF spectra which obscured the weak forward H₂($v' = 1$) scattering signals in that region.

Given that the theoretical simulation is accurate and capable of reproducing the dramatic forward scattering peak, we are now confident of using theory to determine the physical origin of this feature. From the scattering calculation, we computed the state- and angle-resolved integrated opacity function, that is, $d\sigma_{\text{max}}(0, 0 \rightarrow 0, 0; \theta = 0)/d\Omega$ versus J_{max} , where the DCS is the partial sum of waves with $J \leq J_{\text{max}}$, J being the total angular momentum, σ the total reaction cross-section, and Ω the solid angle. This analysis shows that the forward peak in the H₂($v' = 0$) product for $E_c = 1.200$ eV is dominated by the contribution from only a few partial waves near $J = 25$, while the forward peak for the H₂($v' = 1$) product comes mainly from several partial waves near $J = 21$. The narrow-ranged, high- J collisions responsible for the forward scattering implies that the reaction intermediate is formed at very high impact parameter and should rotate quite rapidly. Theoretical results also indicate that other scattering angles show contributions from broad-ranged J collisions, implying that the dynamics for the forward peak is unusual.

To further probe the dynamics associated with the forward peak, a time-delay analysis was carried out using the concept of the angle-resolved scattering time delay proposed in ref. 21. In Fig. 2, we show the time-delay quantity, τ , summed over the final rotational states for the H₂($v' = 0$) product, versus scattering angle. An extra 20 fs of time delay in the forward direction is clearly seen compared to other scattering angles. This shows that the forward-scattering product

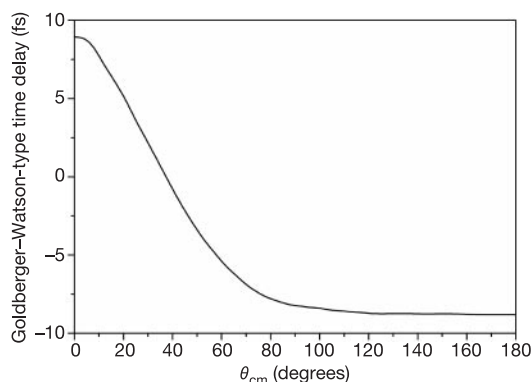


Figure 2 The angle-resolved time delay for H + HD $\rightarrow$ D + H₂($v' = 0$) at $E_c = 1.200$ eV, summed over final rotational states. The angle-resolved time delay is calculated using the method proposed by Goldberger and Watson.²¹ The S-matrix is calculated on a fine grid of energies, so the time-delay quantity can be calculated numerically using the formula:

$$\tau = -\hbar \frac{d}{dE} \text{Arg}(f(\theta, E; \mathbf{n} \rightarrow \mathbf{n}'))$$

where $f(\theta, E; \mathbf{n} \rightarrow \mathbf{n}')$ is the state-to-state scattering amplitude, and $\text{Arg}(f)$ is the phase of the complex function f .

results from a time-delayed mechanism.

From the above analysis, the forward-scattering peak in the H + HD reaction is clearly produced through a rapidly rotating intermediate exhibiting an enhanced time delay. The dynamical picture for this forward peak thus gradually becomes clearer. Further insight into the associated dynamics is provided by a classical trajectory simulation. Trajectories that go to the forward peak show a sideways attack of the H-atom on the HD-diatom (see Fig. 3). At the point where the transition-state region is first reached, the collision complex is already oriented about 70° relative to the centre-of-mass collision axis. Furthermore, the intermediate is formed with a high J ($J \approx 25$) collision and thus rotates rapidly with an angular frequency of $\omega \approx J/I$, where I is the moment of inertia of the intermediate. If the intermediate has a time delay of the order of the lifetime τ , the intermediate can rotate an additional amount $\Delta\theta = \omega\tau$, or about 50° for a 20-fs time delay. Because the intermediate is expected to decay exponentially $e^{-t/\tau}$, there will still be significant probability density at $t = 2.2\tau$ which can take the product into the forward direction. Consistent with this simple picture, we have found that the smallest scattering angles consistently show the longest time delays, both classically and quantum mechanically.

We were interested in the cause of the time delay in the transition-state region for the forward-scattering products. This is complex because the H + HD system is known to support overlapping progressions of quantized bottleneck states corresponding to stretch-bend excitations of the H–H–D triatomic complex^{15,22–25}. Further complicating the analysis is the existence of rotational progressions of states built on each of the vibrationally excited bottleneck states. However, given the preceding discussion, we can focus on the $J = 25$ state as dominantly responsible for the forward peak. Using the centrifugally shifted hamiltonian, $H(J) = H(0) + \hbar^2 \hat{J}^2 / 2\mu R^2$ with R being the H–HD distance, and μ the reduced mass of the two collisional partners, the theoretical spectra of quantized bottleneck states and the wavefunctions of these bottleneck states have been obtained using the spectral quantization technique discussed previously^{23,25}. A quantized bottleneck state is clearly identified at exactly the collisional energy ($E_c = 1.200$ eV) of this experiment (see the peak denoted by an arrow in Fig. 4a).

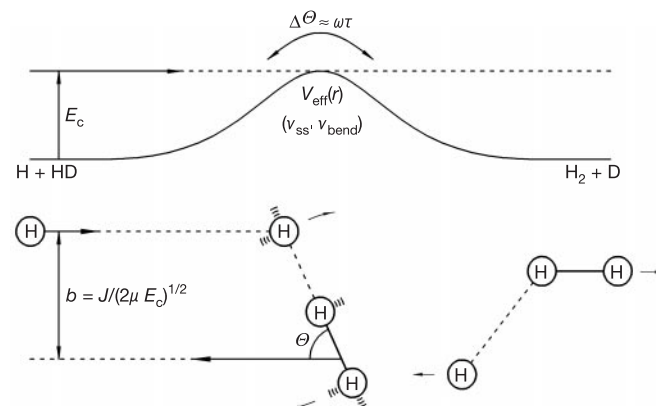


Figure 3 Schematic illustration of the time-delayed reaction mechanism. The scheme gives a picture for the time-delayed mechanism involving a centrifugally shifted quantized bottleneck state, in which the collisional energy of the reaction must be very close to the energy of this state. Analysis of the classical trajectories indicated that the collision complex is already oriented about 70° relative to the centre-of-mass collision axis where the transition state is first reached. The intermediate is formed with a large impact collision ($J \approx 25$) and thus it rotates rapidly with an angular frequency of $\omega \approx J/I$, where I is the moment of inertia of the intermediate. If the intermediate with a sufficient time delay is of the order of the lifetime τ , the intermediate can rotate an additional amount $\Delta\theta = \omega\tau$ to the forward direction before the intermediate falls apart.

In the above analysis, we have shown that the forward scattering in the $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction is clearly produced by a time-delayed mechanism, similar to that for the $\text{H} + \text{D}_2$ reaction^{3,7}. In general, time delay in the transition-state region could be caused by two different mechanisms, that is: (1) by a Feshbach resonance state trapped in an adiabatic potential well as in the $\text{F} + \text{HD} \rightarrow \text{HF} + \text{D}$ reaction; and (2) by a quantized bottleneck state in which time delay results from the slow-down of the motion of the intermediate near

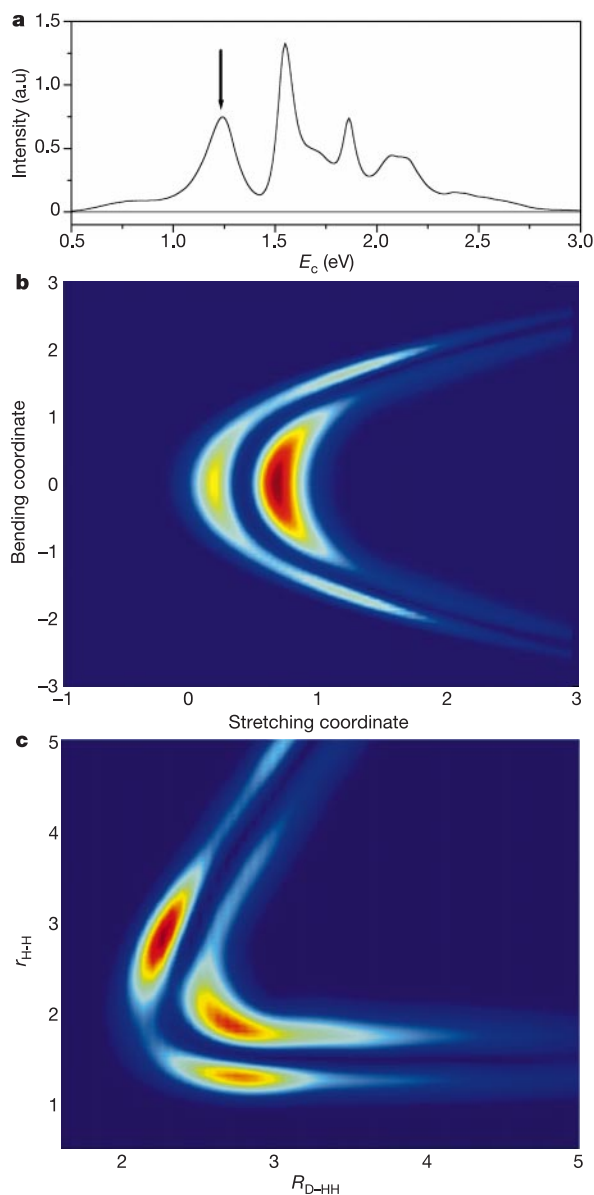


Figure 4 Characteristics of the 'quantized bottleneck state' giving rise to forward product scattering. **a**, The theoretical transition state energy spectrum computed for the $J = 25$ centrifugally shifted hamiltonian. The peak indicated with an arrow at 1.2 eV is assigned to the quantized bottleneck state responsible for the forward scattering. **b**, The probability density of the quantized bottleneck state wavefunction plotted in normal mode coordinates (in units of bohr) obtained using the spectral quantization method. The wavefunction is sliced at the variational transition state and clearly reveals the symmetric stretch and bending excitation. **c**, The probability density of the quantized bottleneck state wavefunction plotted in Jacobi coordinates ($R_{\text{D-HH}}$ versus $r_{\text{H-H}}$ in units of bohr) obtained using the spectral quantization method, where $R_{\text{D-HH}}$ is the Jacobi coordinate between D and H_2 , and $r_{\text{H-H}}$ is the Jacobi coordinate between H and H. Even though the probability density of the wavefunction along the reaction coordinate peaks near the barriers of the adiabatic potential, it is clear that the wavefunction is delocalized with substantial tails extending towards both reactant and product asymptotes.

the top of the adiabatic reaction barrier rather than trapping, during which a few vibrations take place in the coordinates perpendicular to the reaction coordinate. The distinction between these two types of mechanisms and, in fact, the existence and characteristics of quantized bottleneck states have only become clear in recent years^{25,26}. The spectral quantization method shows that no clear Feshbach resonance state is present in the centrifugally shifted vibrationally adiabatic potentials in the $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ system near the energy of the experiment; on the other hand, our results show clearly that the collisional energy ($E_c = 1.200$ eV) in this experiment is very close to the energy of a quantized bottleneck state for the $J = 25$ collisions predominantly responsible for the forward scattering. Therefore, we believe that the time delay observed in the forward scattering could only be produced by the slow-down of the intermediate's motion near the top of the barrier of the specific quantized bottleneck state at about $E_c = 1.2$ eV shown in Fig. 4a. It is clear that this mechanism is significantly different from the Feshbach resonance mechanism in which trapping occurs in an effective potential well. However, there are also similarities between the two mechanisms. In the Feshbach resonance mechanism, the collisional energy (E_c) has to be resonant with the trapped resonance state. Similarly, in the time-delayed mechanism through a quantized bottleneck state, the collisional energy (E_c) has to be close to the energy of the quantized bottleneck state, that is, the barrier height.

The assignment of the quantized bottleneck state responsible for the time-delay is also of interest. The wavefunction of this quantized bottleneck state, obtained by the spectral quantization technique, is shown plotted in symmetric stretch and bend normal-mode coordinates perpendicular to the reaction path in Fig. 4b. There is one node in the symmetric stretch and two nodes along the bending coordinate, corresponding clearly to a quantized bottleneck state with one quantum of symmetric stretch excitation and two quanta of bend. Because there is a 2:1 frequency matching between the symmetric stretch and bending modes, ν_{ss} and ν_{bend} , the quantized bottleneck state characterized should be a mixed state corresponding to the adiabatic barrier states with $(\nu_{\text{ss}}, \nu_{\text{bend}}) = (1, 0)$ and $(0, 2)$. We point out that the wavefunction of the quantized bottleneck state mentioned above is delocalized along the reaction coordinate with long tails extending towards the reactant and product asymptotes (see Fig. 4c), even though the probability density of the wavefunction clearly peaks near the barriers.

Finally, we note that a quantized bottleneck state could essentially affect the reaction dynamics in a similar way (producing forward scattering) to a trapped Feshbach resonance state as in the $\text{F} + \text{HD} \rightarrow \text{HF} + \text{D}$ reaction¹. We also believe that the time-delayed mechanism through a quantized bottleneck state can also explain the forward scattering in other $\text{H} + \text{H}_2$ isotope variant reactions such as the forward scattering observed in the $\text{H} + \text{D}_2$ system³⁻⁵. Such a mechanism should also play an important role in many chemical reactions with barriers at high collisional energies. □

Methods

Crossed-beam scattering using the HRTOF method

The $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction was studied experimentally using the $\text{H}(\text{D})$ atom Rydberg 'tagging' TOF technique developed in refs 11 and 12. The experimental apparatus used in this work is described previously^{27,28}. The excitation of the ground state $\text{D}(1s)$ atom product to its Rydberg state is effected with a two-step excitation scheme. The 121.6-nm vacuum ultraviolet light used in the first-step excitation of the $\text{D}(1s)$ atom to the $2p$ level is generated by a two-photon resonant ($2\omega_1 - \omega_2$) four-wave mixing scheme in the Kr gas cell. Following the first-step vacuum ultraviolet excitation, the $\text{D}(2p)$ atom product is sequentially excited to a higher Rydberg state with the principal quantum number $n \approx 50$ using a 365-nm laser light. These two excitation laser pulses are both overlapped in space and synchronized in time in the detection region. The neutral Rydberg 'tagged' D atoms then fly a certain distance and reach a Z-stack multichannel plate detector with a fine metal grid (grounded) in the front. After passing through the grid, the Rydberg D atom products are then immediately field-ionized by the electric field applied between the front plate of the multichannel plate detector and the fine metal grid, and detected by the multichannel plate detector. Two parallel molecular beams (HI and HD) are generated with similar

pulsed valves in this experiment. The H atom beam is produced by the photolysis of the HI molecule at 266 nm using a Nd:YAG fourth harmonic laser output. There are two different photolysis channels, corresponding to the two spin-orbit states of the iodine atom, which give rise to two sharp peaks of H-atom velocity distribution: a fast component at $17,470 \text{ m s}^{-1}$ and a slow component at $11,230 \text{ m s}^{-1}$. The H beam is then crossed at 90° with the HD molecular beam. The HD beam was produced by an adiabatic expansion through a nozzle cooled to the liquid nitrogen temperature, ensuring that all the HD molecules be in the $j = 0$ level. In this experiment, the fast HI-photolysis channel is used, and the collisional energy is 1.200 eV. The uncertainty in the collisional energy is estimated to be about 15 meV. The D atom products were detected using the Rydberg H atom TOF method with a rotatable multichannel plate detector.

Quantum scattering calculations

The $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ reaction has been investigated theoretically using full quantum scattering calculations. The S-matrix for $\text{H} + \text{HD}$ was calculated using quantum-reactive-scattering methods described previously¹⁸, on the BKM2 PES²⁰. We used the time-independent coupled-channel hyperspherical coordinate method developed in ref. 17 to compute the scattering amplitudes for all open channels for the three final particle arrangements. The convergence parameters chosen were similar to those we used previously¹⁸ for $\text{H} + \text{D}_2$ except that the maximum allowed helicity quantum number was taken as $k_{\text{max}} = 6$ to ensure convergence. We have carried out scattering calculations on a grid of 56 total energies at increments of 0.023 eV for $E_{\text{tot}} = 0.5\text{--}1.76 \text{ eV}$. The calculations were performed for total angular momentum quantum number J up to $J_{\text{max}} = 35$ in order to give well-converged ICS and DCS in the energy range. The ICS and DCS were computed from the S-matrix using standard formulae where the helicities were summed over the final asymptotic states. To simulate the molecular beam experiments, only the lowest reactant state, $\text{HD}(\nu = j = 0)$, needs to be considered. In addition, the effect of spin statistics in the final H_2 -diatom has also been considered. An S-matrix is computed, assuming that each nucleus is distinguishable. To enforce the correct Fermi–Dirac statistics on the final product states, the post-antisymmetrization procedure^{13,29} is then employed in computing the final-state-specific DCS.

Received 6 June; accepted 9 August 2002; doi:10.1038/nature01068.

- Skodje, R. T. *et al.* Resonance-mediated chemical reaction: $\text{F} + \text{HD} \rightarrow \text{HF} + \text{D}$. *Phys. Rev. Lett.* **85**, 1206–1209 (2000).
- Neumark, D. M., Wodtke, A. M., Robinson, G. N., Hayden, C. C. & Lee, Y. T. Experimental investigation of resonances in reactive scattering: The $\text{F} + \text{H}_2$ reaction. *Phys. Rev. Lett.* **53**, 226–229 (1984).
- Fernandes-Alonso, F. *et al.* Evidence for scattering resonances in the $\text{H} + \text{D}_2$ reaction. *Angew. Chem. Int. Edn Engl.* **39**, 2748–2752 (2001).
- Fernandez-Alonso, F. *et al.* Forward scattering in the $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$ reaction: comparison between experiment and theoretical predictions. *J. Chem. Phys.* **115**, 4534–4545 (2001).
- Althorpe, S. C. *et al.* Observation and interpretation of a time-delayed mechanism in the hydrogen exchange reaction. *Nature* **416**, 67–70 (2002).
- Aoiz, F. J., Herrero, V. J. & Saez Rabanos, V. Quasiclassical state to state reaction cross sections for $\text{D} + \text{H}_2(\nu = 0, j = 0) \rightarrow \text{HD}(\nu', j') + \text{H}$. Formation and characteristics of short-lived collision complexes. *J. Chem. Phys.* **97**, 7423–7436 (1992).
- Allison, T. C., Friedman, R. S., Kaufman, D. J. & Truhlar, D. G. Analysis of the resonance in $\text{H} + \text{D}_2 \rightarrow \text{HD}(\nu' = 3) + \text{D}$. *Chem. Phys. Lett.* **327**, 439–445 (2000).
- Buntin, S. A., Giese, C. F. & Gentry, W. R. State-resolved differential cross sections for the reaction $\text{D} + \text{H}_2 \rightarrow \text{HD} + \text{H}$. *J. Chem. Phys.* **87**, 1443–1445 (1987).
- Kliner, D. A., Adelman, D. E. & Zare, R. N. Comparison of experimental and theoretical integral cross sections for $\text{D} + \text{H}_2(\nu = 1, j = 1) \rightarrow \text{HD}(\nu' = 1, j') + \text{H}$. *J. Chem. Phys.* **95**, 1648–1662 (1991).
- Kitopoulos, T. N., Buntine, M. A., Balwin, D. P., Zare, R. N. & Chandler, D. W. Reaction product imaging: The $\text{H} + \text{D}_2$ reaction. *Science* **260**, 1605–1610 (1993).
- Schneider, L. *et al.* Experimental studies and theoretical predictions for the $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$ reaction. *Science* **269**, 207–210 (1995).
- Schneider, L., Seekamp-Rahn, K., Wrede, E. & Welge, K. H. Experimental determination of quantum state resolved differential cross sections for the hydrogen exchange reaction $\text{H} + \text{D}_2 \rightarrow \text{HD} + \text{D}$. *J. Chem. Phys.* **107**, 6175–6195 (1997).
- Zhang, J. Z. H. & Miller, W. H. Quantum reactive scattering via the S-matrix version of the Kohn variational principle: Differential and integral cross sections for $\text{D} + \text{H}_2 \rightarrow \text{HD} + \text{H}$. *J. Chem. Phys.* **91**, 1528–1547 (1991).
- Zhao, M., Truhlar, D. G., Schwenke, D. W. & Kouri, D. J. Effect of rotational excitation on state-to-state differential cross sections: deuterium atom + hydrogen $\rightarrow$ hydrogen deuteride + hydrogen atom. *J. Phys. Chem.* **94**, 7074–7090 (1990).
- Zhao, M. *et al.* Spectroscopic analysis of transition state energy levels: Bending–rotational spectrum and lifetime analysis of H_3 quasisubstituted states. *J. Chem. Phys.* **91**, 5302–5309 (1989).
- D'Mello, M. J., Manolopoulos, D. E. & Wyatt, R. E. Quantum dynamics of the $\text{H} + \text{D}_2 \rightarrow \text{D} + \text{HD}$ reaction: Comparison with experiment. *J. Chem. Phys.* **94**, 5985–5993 (1991).
- Skouteris, D. M., Castillo, J. F. & Manolopoulos, D. E. ABC: a quantum reactive scattering program. *Comput. Phys. Commun.* **133**, 128–135 (2000).
- Chao, S. D. & Skodje, R. T. The search for resonance signatures in $\text{H} + \text{D}_2$ reaction dynamics. *Chem. Phys. Lett.* **336**, 364–370 (2001).
- Schneider, L. *et al.* Photodissociation dynamics of H_2S at 121.6 nm and a determination of the potential energy function of $\text{SH}(A^2\Sigma^+)$. *J. Chem. Phys.* **92**, 7027–7037 (1990).
- Boothroyd, A. I., Keogh, W. J., Martin, P. G. & Peterson, M. R. A refined H_3 potential energy surface. *J. Chem. Phys.* **104**, 7139–7152 (1996).
- Goldberger, M. L. & Watson, K. M. *Collision Theory* (Wiley, New York, 1964).
- Cuccaro, S. A., Hipes, P. G. & Kuppermann, A. Symmetry analysis of accurate $\text{H} + \text{H}_2$ resonances for low partial waves. *Chem. Phys. Lett.* **157**, 440–446 (1989).
- Skodje, R. T., Sadeghi, R., Koppel, H. & Krause, J. L. Spectral quantization of transition state dynamics for the three-dimensional $\text{H} + \text{H}_2$ reaction. *J. Chem. Phys.* **101**, 1725–1729 (1994).
- Chatfield, D. C., Mielke, S. L., Allison, T. C. & Truhlar, D. G. Quantized dynamical bottlenecks and

transition state control of the reaction of D with H_2 : Effect of varying the total angular momentum. *J. Chem. Phys.* **112**, 8387–8408 (2000).

- Sadeghi, R. & Skodje, R. T. Barriers, thresholds, and resonances: Spectral quantization of the transition state for the collinear $\text{D} + \text{H}_2$ reaction. *J. Chem. Phys.* **102**, 193–213 (1995).
- Chatfield, D. C., Friedman, R. S., Schwenke, D. W. & Truhlar, D. G. Control of chemical reactivity by quantized transition states. *J. Phys. Chem.* **96**, 2414–2421 (1992).
- Harich, S. A., Dai, D., Yang, X., Chao, S. D. & Skodje, R. T. State-to-state dynamics of $\text{H} + \text{HD} \rightarrow \text{H}_2 + \text{D}$ at 0.5 eV: A combined theoretical and experimental study. *J. Chem. Phys.* **116**, 4769–4772 (2002).
- Liu, X., Lin, J. J., Harich, S. A., Schatz, G. C. & Yang, X. A quantum state-resolved insertion reaction: $\text{O}(^1\text{D}) + \text{H}_2(j = 0) \rightarrow \text{OH}(^2\Pi, \nu, N) + \text{H}(^2\text{S})$. *Science* **289**, 1536–1538 (2000).
- Kuppermann, A., Schatz, G. C. & Baer, M. Quantum mechanical reactive scattering for planar atom plus diatom systems. I. Theory. *J. Chem. Phys.* **65**, 4596–4623 (1976).

Acknowledgements

We thank K. Liu and Y.T. Lee for many helpful discussions. The experimental work was supported mainly by the National Science Council and Academia Sinica of Taiwan, and the theoretical effort was supported by the National Science Foundation of USA, and by the Ministry of Education, Culture, Sports, Science and Technology of Japan. D.D. and X.Y. also acknowledge support for this work at DICP by the Ministry of Science & Technology of China.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.Y. (e-mail: xmyang@po.iam.s.sinica.edu.tw) or R.T.S. (e-mail: skodje@spot.colorado.edu).

An all-organic composite actuator material with a high dielectric constant

Q. M. Zhang, Hengfeng Li, Martin Poh, Feng Xia, Z.-Y. Cheng, Haisheng Xu & Cheng Huang

Materials Research Institute and Electrical Engineering Department, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Electroactive polymers (EAPs) can behave as actuators, changing their shape in response to electrical stimulation. EAPs that are controlled by external electric fields—referred to here as field-type EAPs—include ferroelectric polymers, electrostrictive polymers, dielectric elastomers and liquid crystal polymers^{1–6}. Field-type EAPs can exhibit fast response speeds, low hysteresis^{1–8} and strain levels far above those of traditional piezoelectric materials^{4,6,9,10}, with elastic energy densities even higher than those of piezoceramics^{4,5,9–11}. However, these polymers also require a high field ($>70 \text{ V } \mu\text{m}^{-1}$) to generate such high elastic energy densities ($>0.1 \text{ J cm}^{-3}$; refs 4, 5, 9, 10). Here we report a new class of all-organic field-type EAP composites, which can exhibit high elastic energy densities induced by an electric field of only $13 \text{ V } \mu\text{m}^{-1}$. The composites are fabricated from an organic filler material possessing very high dielectric constant dispersed in an electrostrictive polymer matrix. The composites can exhibit high net dielectric constants while retaining the flexibility of the matrix. These all-organic actuators could find applications as artificial muscles, ‘smart skins’ for drug reduction, and in microfluidic systems for drug delivery^{1–3,12}.

The elastic energy density is a key parameter, measuring both the stress and strain generation capability of an actuator material. The large operation field required to generate high elastic energy densities in EAPs follows from the principle of energy conservation. To illustrate this point, we take a field-type EAP which is assumed to be a linear dielectric and elastic material. The stored elastic energy density U_s when a polymer is strained is $U_s = 1/2 Y S^2$, where Y is

the Young's modulus and S is the strain. For a field-type EAP, the total elastic energy density from all the strains generated cannot exceed the input electric energy density because energy must be conserved. As a linear dielectric material, this input electric energy density is $U_E = 1/2 K \epsilon_0 E^2$, where E is the applied field, ϵ_0 is the vacuum dielectric permittivity ($= 8.85 \times 10^{-12} \text{ F m}^{-1}$), and K is the dielectric constant of the polymer. In most of the polymeric materials, the dielectric constant K is less than ten¹³ which is far below those in the inorganic materials, many of which can reach more than 5,000 (ref. 14). So in order to generate a high input electric energy density which can be converted to strain energy, a high electric field is required. For example, to generate a U_s of 0.1 J cm^{-3} —which corresponds to the elastic energy density of the best-performance piezoelectric and electrostrictive ceramics^{14,15}—in a polymer with a dielectric constant of ten, and assuming a 50% energy conversion efficiency, which is very high for the current field-type EAPs, the field required is $67 \text{ V } \mu\text{m}^{-1}$. Therefore, in order to reduce the applied field substantially in the field-type EAPs while retaining the high elastic energy density that is required in many practical applications, we need to raise the dielectric constant of this class of polymers substantially.

The composite approach—where high-dielectric-constant particulates were added to a polymer matrix—has previously been used to raise the dielectric constant of polymer-based materials^{3,16}. However, because these high-dielectric-constant fillers (frequently ceramics) also possess an elastic modulus that is much higher than that of the polymers, the resulting composite also showed an elastic modulus much higher than that of the polymer matrix and thus a loss of flexibility. In addition, the low dielectric constant of the polymer matrix (≤ 10) also resulted in the composite dielectric constant still falling below 100.

Recent investigations^{4,9} have shown that the electrostrictive poly(vinylidene fluoride-trifluoroethylene) P(VDF-TrFE), belonging to the class of relaxor ferroelectrics⁴, a class of ferroelectric material which in many ways resembles polarglass, exhibits a relatively high room-temperature dielectric constant (>40) and a high electrostriction (strain $>4\%$). Both characteristics are highly desirable in a composite to achieve a high dielectric constant and high field-induced strain, and so we chose the electrostrictive P(VDF-TrFE) as the matrix material for our composite.

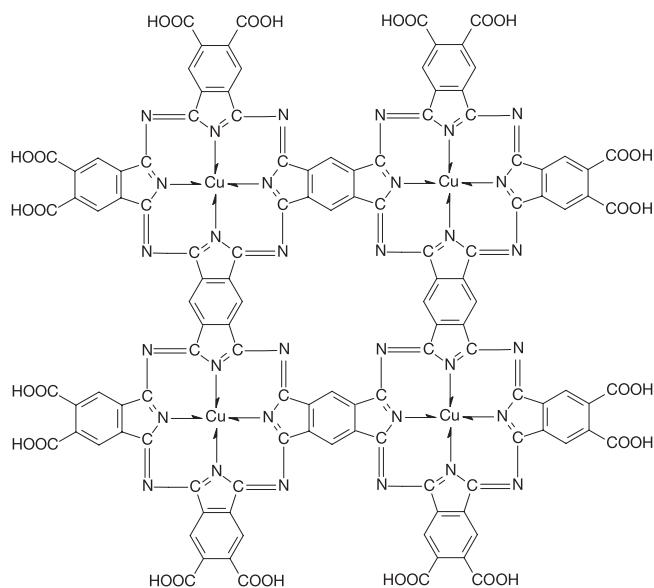


Figure 1 Schematic of the copper-phthalocyanine (CuPc) oligomer used as the high-dielectric-constant filler.

We selected a metallophthalocyanine (MtPc) oligomer, copper-phthalocyanine (CuPc), as the filler of high dielectric constant ($>10,000$). Molecular solids formed from MtPc are a class of organic semiconductor materials that are finding applications in organic transistors, chemical sensors, and organic electroluminescent devices^{17,18}. A dielectric constant as high as 10^5 has been observed in CuPc oligomers, whose molecular structure is shown schematically in Fig. 1 (refs 19, 20). The large dielectric constant can be explained in terms of the electron delocalization within MtPc molecules^{20,21}. The easy displacement of the electrons under electric fields from the conjugated π -bonds within the entire molecule results in a high dielectric response. Furthermore, the weak van der Waals intermolecular forces render the molecular solids formed with an elastic modulus not much higher than the polymer matrix. On the other hand, MtPc solids are difficult to process and show a high dielectric loss owing to the long-range intermolecular hopping of electrons²². Therefore, the polymer matrix also forms insulation layers to reduce significantly the dielectric loss of the filler.

The resulting composite exhibits almost the same elastic modulus as the polymer matrix and retains its flexibility. For composites containing 40 wt% to 55 wt% of CuPc, which exhibit high dielectric constant while the dielectric loss remains low, the elastic modulus at room temperature is in the range of 0.6 GPa to 1.2 GPa.

The field-induced strain was measured at different applied fields at room temperature and under an a.c. field of 1 Hz. Presented in Fig. 2 is the strain amplitude measured versus the applied-field amplitude for the composite containing 40 wt% of CuPc. A strain near -2% (-1.91%), which is comparable to those in the electrostrictive P(VDF-TrFE)-based polymers, can be induced under a field of $13 \text{ V } \mu\text{m}^{-1}$. Combining the strain and elastic modulus (0.75 GPa) data yields the elastic energy density of 0.13 J cm^{-3} for the composite under a field of $13 \text{ V } \mu\text{m}^{-1}$. The results indicate that this all-organic composite approach can result in a more than six times reduction in the applied field compared with the other field-type EAPs with similar strain and elastic energy density^{4–6,9,10}. For comparison, the induced strain from the electrostrictive P(VDF-TrFE) copolymer in the same field range is also presented⁹. For the current composite, the applied field of $13 \text{ V } \mu\text{m}^{-1}$ corresponds to the breakdown field of the material (the highest field that can be applied to the material without electrical shorting), which can be increased by improving the composite quality through further investigation on composite fabrication methods. At fields lower than $10 \text{ V } \mu\text{m}^{-1}$, the composite can be operated over several hours without changes in the strain response.

The results of the strain measurement indicate that the field-induced strain S is proportional to the square of the applied electric

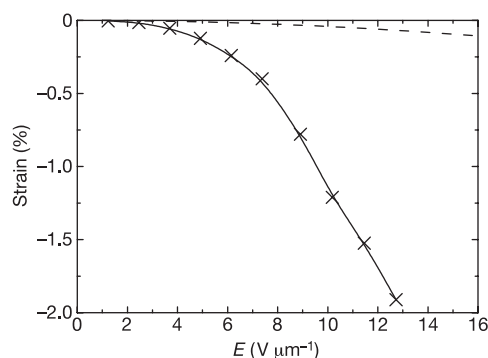


Figure 2 The strain amplitude as a function of the applied-field amplitude measured at room temperature. The applied-field frequency is 1 Hz. Crosses are the data and the solid curve is a guide to the eye. For comparison, the strain from the electrostrictive P(VDF-TrFE) copolymer at the same field range is also shown (the dashed curve).

field, suggesting that the strain response originates from either Maxwell stress (the electrostatic force), or the electrostriction, or both^{4,5,23,24}. When a dielectric material is subject to an electric field E , it will experience an electrostatic force (Maxwell stress) that is due to the Coulomb force between charges. The strain induced by the Maxwell stress along the applied-field direction is²³:

$$S = 1/2K\epsilon_0 E^2(1 + 2\sigma)/Y \quad (1)$$

Here we assumed that the dielectric material is isotropic and σ is the Poisson's ratio. The electrostriction is strain-generated owing to a change in the polarization in a material under constant stress, which is equal to $S = ME^2$, where M is the electrostrictive coefficient and is proportional to the square of the dielectric constant²⁴.

Figure 3a presents the dielectric constant as a function of the applied-field amplitude at room temperature and 1 Hz for the composite containing 40 wt% CuPc. At low fields, the composite has a dielectric constant of 225 and a loss factor ($D = K''/K'$) of 0.4. With increased field amplitude, the dielectric constant increases and at $13 \text{ V } \mu\text{m}^{-1}$, the dielectric constant is about 425 and the loss

remains relatively low (~ 0.7) compared with that of the polymer matrix (the resistivity of the composite is higher than $10^8 \Omega \text{ m}$). For CuPc, the dielectric constants should not increase with field amplitude²¹. The observed increase of the dielectric constant with the applied-field amplitude is from the P(VDF-TrFE) matrix, which we also measured (Fig. 3b). Such an increase in the dielectric constant with the applied-field amplitude in a ferroelectric material is believed to be caused mainly by the interaction of the polarization with the defect fields in the materials²⁵.

From the dielectric and elastic data, the contributions to the strain response from the Maxwell stress and electrostriction may be estimated. Assuming that the composite is homogeneous, the strain due to Maxwell stress under a field of $13 \text{ V } \mu\text{m}^{-1}$ is about -0.1% . For the electrostriction from the polymer matrix, assuming that all the applied field is totally loaded onto the polymer matrix because its dielectric constant is much lower than that of CuPc, the strain is also about -0.1% . The measured strain response, therefore, is about one order of magnitude higher than the combined strain from the two.

In an early study on the field-induced strain in a polyurethane elastomer, it was observed that a nonuniform electric-field distribution can significantly enhance the strain response (more than five times in that case) if the strain is proportional to the square of the local field²⁶. The composites investigated here are highly heterogeneous, so that a large variation in the local fields is likely, which will enhance the strain response. In addition, bond length and conformation changes in CuPc due to electron motions, as well as the CuPc molecular reorientation under external fields, may also contribute to the strain response²⁷.

Figure 3c shows the frequency dependence of the weak-field dielectric constant of the composite. The strong frequency dispersion is due to the space charge polarization (delocalized electrons) in CuPc, which will also result in frequency dependence of the strain response in the composite. For instance, the strain measured at 0.1 Hz is about 1.6 times higher than that measured at 1 Hz.

We note that the approach of using the delocalized charge phenomenon in a composite in which insulation layers block the long-range space charge conduction to realize very high dielectric constant is analogous to that in the ceramic capacitor where semiconductor cores and insulation boundary-layers form the so-called internal boundary layer capacitor, resulting in a material with dielectric constants exceeding 50,000 (refs 14, 28). With further improvement in the composite fabrication process, including using nanoparticulates of CuPc and coating the CuPc particles completely with insulation layers, both the dielectric properties and the breakdown field can be increased substantially. In view of the availability of a large number of organic solids with super-dielectric constant, the results presented here demonstrate the potential of this all-organic composite approach in generating high strain and high elastic energy density under low applied fields in polymer-like materials²¹. $\square$

Methods

The copper-phthalocyanine (CuPc) oligomer was synthesized by the solution method²⁹. Copper sulphate pentahydrate, pyromellitic dianhydride, urea, ammonium chloride, and ammonium molybdate were ground together and then placed in a three-necked flask containing nitrobenzene. The temperature of the flask was maintained at 185°C for 12 h. The obtained solid material was boiled with hydrochloric acid, and then treated with potassium hydroxide. The final powder was dried at room temperature under vacuum. The typical particle size thus obtained, as measured by scanning electron microscopy, was below $1 \mu\text{m}$.

The P(VDF-TrFE) 50/50 mol% copolymer (purchased from Solvay & Cie.) was used as the matrix, which can be easily converted to the electrostrictive material by high energy electron irradiation^{4,9}. Composite films were prepared using the solution cast method. P(VDF-TrFE) copolymer was dissolved in dimethylformamide (DMF) and then a proper amount of CuPc powder (determined by the wt% of CuPc in the composite) was added to the solution. The solution was ultrasonically stirred to disperse the CuPc powder. After that, the solution was poured onto a glass plate and dried at 70°C . The typical film thickness was about $40 \mu\text{m}$. The composites were irradiated at 100°C with electrons of 1.2 MeV energy to convert the copolymer matrix into an electrostrictive polymer with an

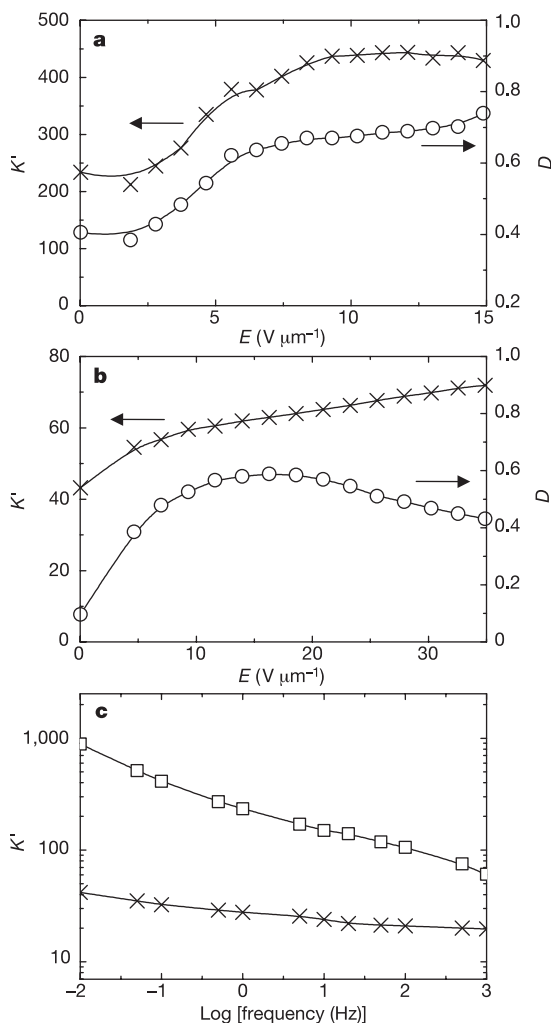


Figure 3 Dielectric properties of the all-organic composites. **a, b**, The real part of the dielectric constant (K') and dielectric loss (D) as a function of the applied-field amplitude for the composite with 40 wt% CuPc (**a**) and the polymer matrix (**b**). The field frequency is 1 Hz. **c**, The weak-field (100 V cm^{-1}) dielectric constant as a function of frequency for the composite (squares) and the polymer matrix (crosses). Data points are shown and solid curves are a guide to the eye.

improved room-temperature dielectric constant.

The dielectric properties of the composites under different applied-field strength were characterized by directly measuring the magnitude and the phase of the current passing through the composite under a given a.c. voltage. From the complex electric impedance $Z^* = 1/(j\omega C^*)$ and $C^* = (K' - jK'')\epsilon_0 A/t$, both the real and imaginary parts (K' and K'') of the dielectric constant can be determined, where A is the area and t is the thickness of the capacitor, and ω is the angular frequency. The elastic modulus was determined using a commercial dynamic mechanical analyser (TA Instruments, DMA2980). The electric-field-induced strain along the applied-field direction (longitudinal strain) was measured using a piezobimorph-based sensor³⁰.

Received 18 April; accepted 25 July 2002; doi:10.1038/nature01021.

1. Bar-Cohen, Y. (ed.) *Electroactive Polymer Actuators as Artificial Muscles* (SPIE, Bellingham, WA, 2001).
2. Zhang, Q. M., Furukawa, T., Bar-Cohen, Y. & Scheinbeim, J. (eds) *Electroactive Polymers* (MRS Symp. Proc. Vol. 600, MRS, Warrendale PA, 1999).
3. Nalwa, H. (ed.) *Ferroelectric Polymers* (Marcel Dekker, New York, 1995).
4. Zhang, Q. M., Bharti, V. & Zhao, X. Giant electrostriction and relaxor ferroelectric behavior in electron-irradiated poly(vinylidene fluoride-trifluoroethylene) copolymer. *Science* **280**, 2101–2104 (1998).
5. Pelrine, R., Kornbluh, R., Pei, Q. & Joseph, J. Highspeed electrically actuated elastomers with strain greater than 100%. *Science* **287**, 836–839 (2000).
6. Lehmann, W. et al. Giant lateral electrostriction in ferroelectric liquid-crystalline elastomers. *Nature* **410**, 447–450 (2001).
7. Baughman, R. H. et al. Carbon nanotube actuators. *Science* **284**, 1340–1344 (1999).
8. Osada, Y., Okuzaki, H. & Hori, H. A polymer gel with electrically driven motility. *Nature* **355**, 242–244 (1992).
9. Cheng, Z.-Y. et al. Electrostrictive poly(vinylidene fluoride-trifluoroethylene) copolymers. *Sens. Actuators A* **90**, 138–147 (2001).
10. Xu, H. et al. Ferroelectric and electromechanical properties of poly(vinylidene-fluoride-trifluoroethylene-chlorotrifluoroethylene) terpolymer. *Appl. Phys. Lett.* **78**, 2360–2362 (2001).
11. Huber, J. E., Fleck, N. A. & Ashby, M. F. The selection of mechanical actuators based on performance indices. *Proc. R. Soc. Lond. A* **453**, 2185–2205 (1997).
12. Dario, P., Carrozza, M., Benvenuto, A. & Mencias, A. Micro-systems in biomedical applications. *J. Micromech. Microeng.* **10**, 235–244 (2000).
13. McCrum, N., Read, B. E. & Williams, G. *Anelastic and Dielectric Effects in Polymeric Solids* (Dover, New York, 1967).
14. Moulson, A. & Herbert, J. *Electroceramics* (Chapman & Hall, London, 1995).
15. Cross, L. E. Ferroelectric ceramics: materials and application issues. *Ceram. Trans.* **68**, 15–55 (1996).
16. Safari, A., Sa-gong, G., Giniwicz, J. & Newnham, R. Composite piezoelectric sensors. *Proc. 21st Univ. Conf. Ceram. Sci.* **20**, 445–455 (1986).
17. Bao, Z., Lovinger, A. & Dodabalapur, A. Highly ordered vacuum-deposited thin films of metallophthalocyanines and their application in field-effect transistors. *Adv. Mater.* **9**, 42–45 (1997).
18. Tominaga, T., Hayashi, K. & Toshima, N. Accelerated hole transfer by double-layered metallophthalocyanine thin film for effective electroluminescence. *Appl. Phys. Lett.* **70**, 762–763 (1997).
19. Nalwa, H. S., Dalton, L. & Vasudevan, P. Dielectric properties of copper-phthalocyanine polymer. *Eur. Polym. J.* **21**, 943–947 (1985).
20. Phougat, N., Vasudevan, P. & Nalwa, H. *Handbook of Low and High Dielectric Constant of Materials and Their Applications* Ch. 8 (ed. Nalwa, H.) (Academic, London, 1999).
21. Vijayakumar, P. & Pohl, H. A. Giant polarization in stable polymeric dielectrics. *J. Polym. Sci. Polym. Phys. Ed.* **22**, 1439–1451 (1984).
22. Gould, P. D. Structure and electrical conduction properties of phthalocyanine thin films. *Coord. Chem. Rev.* **156**, 237–274 (1996).
23. Landau, L. D. & Lifshitz, E. M. *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1970).
24. Newnham, R., Sundar, V., Yimmirun, R., Su, J. & Zhang, Q. M. Electrostriction in dielectric materials. *Ceram. Trans.* **88**, 15–39 (1998).
25. Damjanovic, D. Logarithmic frequency dependence of the piezoelectric effect due to pinning of ferroelectric-ferroelastic domain walls. *Phys. Rev. B* **55**, R649–R652 (1997).
26. Su, J., Zhang, Q. M. & Ting, R. Y. Space-charge-enhanced electromechanical response in thin-film polyurethane elastomers. *Appl. Phys. Lett.* **71**, 386–388 (1997).
27. Hayashi, K., Kawato, S., Fujii, Y., Horiuchi, T. & Matsushige, K. Effect of applied electric field on the molecular orientation of epitaxially grown organic films. *Appl. Phys. Lett.* **70**, 1384–1386 (1997).
28. Ai-Aliak, H., Illingsworth, J., Brinkman, A., Russell, G. J. & Woods, J. The effects of donor dopant concentration on the grain boundary layer characteristics in *n*-doped BaTiO₃ ceramics. *J. Appl. Phys.* **64**, 6477–6482 (1998).
29. Achar, B. N., Fohlen, G. G. & Parker, J. A. Phthalocyanine polymers. II Synthesis and characterization of some metal phthalocyanine sheet oligomers. *J. Polym. Sci. Polym. Chem. Ed.* **20**, 1785–1790 (1982).
30. Su, J., Moses, P. & Zhang, Q. M. A piezoelectric bimorph based dilatometer for field induced strain measurement in soft and thin free standing polymer films. *Rev. Sci. Instrum.* **69**, 2480–2484 (1998).

Acknowledgements

This work was supported by the National Institutes of Health, the Office of Naval Research, and Defense Advanced Research Projects Agency.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Q.M.Z. (e-mail: qxz1@psu.edu).

Testing time-predictable earthquake recurrence by direct measurement of strain accumulation and release

Jessica Murray & Paul Segall

Geophysics Department, Stanford University, Stanford, California 94305-2215, USA

Probabilistic estimates of earthquake hazard use various models for the temporal distribution of earthquakes, including the ‘time-predictable’ recurrence model formulated by Shimazaki and Nakata¹ (which incorporates the concept of elastic rebound described as early as 1910 by H. F. Reid²). This model states that an earthquake occurs when the fault recovers the stress relieved in the most recent earthquake. Unlike time-independent models (for example, Poisson probability), the time-predictable model is thought to encompass some of the physics behind the earthquake cycle, in that earthquake probability increases with time. The time-predictable model is therefore often preferred when adequate data are available, and it is incorporated in hazard predictions for many earthquake-prone regions, including northern California³, southern California^{4,5}, New Zealand⁶ and Japan⁷. Here we show that the model fails in what should be an ideal locale for its application — Parkfield, California. We estimate rigorous bounds on the predicted recurrence time of the magnitude ~6 1966 Parkfield earthquake through inversion of geodetic measurements and we show that, according to the time-predictable model, another earthquake should have occurred by 1987. The model’s poor performance in a relatively simple tectonic setting does not bode well for its successful application to the many areas of the world characterized by complex fault interactions.

The time-predictable model states that the time to the next earthquake is the stress drop in the most recent earthquake divided by the fault stressing-rate. Although stress is difficult to measure *in situ*, the Earth’s crust is linearly elastic, so strain is proportional to stress⁸. Geodetic measurements may be used to estimate both the strain released in an earthquake and the interseismic strain rate. If we can use such estimates to show with a high degree of confidence that the strain released in the most recent earthquake has already accumulated without a subsequent earthquake, then the time-predictable model has failed.

The Parkfield segment of the San Andreas fault (Fig. 1a) is a natural choice for a test of the time-predictable model. At least five historic earthquakes of $M \approx 6$ have occurred in this area, most recently in 1966. The average time between earthquakes is about 22 years. The three latest events had striking similarities⁹, including comparable seismic moment and epicentral location. Inferred epicentres for the older events, although uncertain, are consistent with this location. The absence of active sub-parallel faults in this section of the San Andreas system makes Parkfield a particularly good place to test the time-predictable model. This test is possible because geodetic data have been collected here for over 40 yr (refs 10–12), making Parkfield one of the few places where strain has been measured for times exceeding the average inter-event time.

Inversion of geodetic data collected since 1966 has shown that although the San Andreas exhibits shallow creep near Parkfield, the fault is either locked or slowly slipping at depth^{12–14} (see Fig. 1b). This demonstrates that strain has been accumulating throughout what has now become the longest recorded interseismic period for $M \approx 6$ earthquakes at Parkfield.

The coseismic moment, a measure of strain release, is given by

$$M_o = \mu \iint_A s dA \quad (1)$$

where μ is the shear modulus, A is the area that slipped, and s is the (spatially variable) amount of slip. We approximate the complex interseismic loading of the seismogenic zone by aseismic slip at a constant rate on a downward extension of the fault¹⁵. If the seismogenic fault were slipping everywhere at this long-term deep slip-rate, no strain would accumulate. With a spatially variable slip-rate, however, some parts of the fault experience a slip deficit relative to the long-term slip-rate. The moment deficit rate is

$$\dot{M}_d = \mu \iint_A (\dot{s}_\infty - \dot{s}) dA \quad (2)$$

where $\dot{s}_\infty$ is the long-term deep slip-rate, $\dot{s}$ the (spatially variable) slip-rate, and $\dot{s}_\infty - \dot{s}$ the slip deficit rate. The time-predictable inter-event time, t_i , can be written as:

$$t_i = M_o / \dot{M}_d \quad (3)$$

Given certain assumptions about the fault stress⁸, t_i measures the

time required for elastic strain energy to recover in the crustal volume surrounding the fault.

We used two geodetic data sets: (1) trilateration measurements of the distances between benchmarks¹¹; and (2) global positioning system (GPS) estimates of benchmark positions¹². Trilateration measurements repeated over time can be used to estimate rates of distance change, whereas GPS measurements over time produce velocities of individual stations (Fig. 1a). We inverted these data for the slip distribution of the 1966 earthquake and the slip-rate distribution for the interseismic period (1966 to 1998), from which the coseismic moment (M_o) and interseismic moment deficit rate ($\dot{M}_d$) can be derived.

These inversions require the choice of a model fault geometry (Fig. 1), one component of which is a transition depth below which

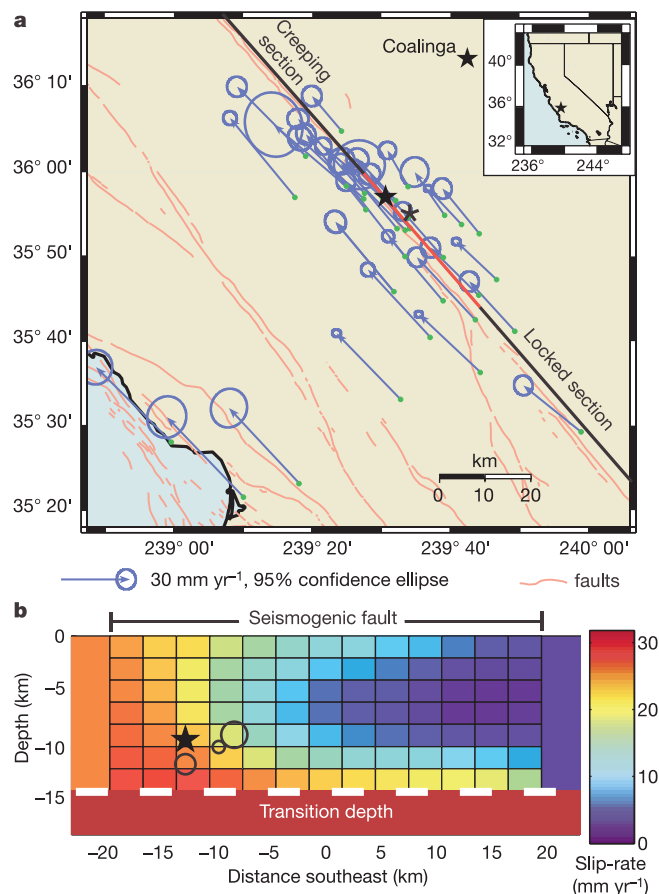


Figure 1 Parkfield segment of the San Andreas fault. The unlabelled stars on **a** and **b** represent the 1966 earthquake. **a**, The San Andreas fault near the town of Parkfield (asterisk) forms a transition between the creeping segment to the northwest and the locked section to the southeast. Vectors indicate GPS velocities (1991–1998) relative to North America. Black and red line is model fault trace; red portion is the seismogenic fault which is divided into subfaults (dislocations). **b**, Slip-rate estimated for each subfault after Murray *et al.*¹² Zero on x-axis is the midpoint of seismogenic fault in **a**. The northwestern endpoint is at 36° N, 239° 27' E; the southeastern endpoint is at 35° 44' N, 239° 44' E. Circles are earthquakes occurring in (from northwest to southeast) November 1993 ($M = 4.6$), October 1992 ($M = 4.3$) and December 1994 ($M = 4.7$). Area of circles is equivalent to the rupture area of a 3 MPa stress drop crack³⁰.

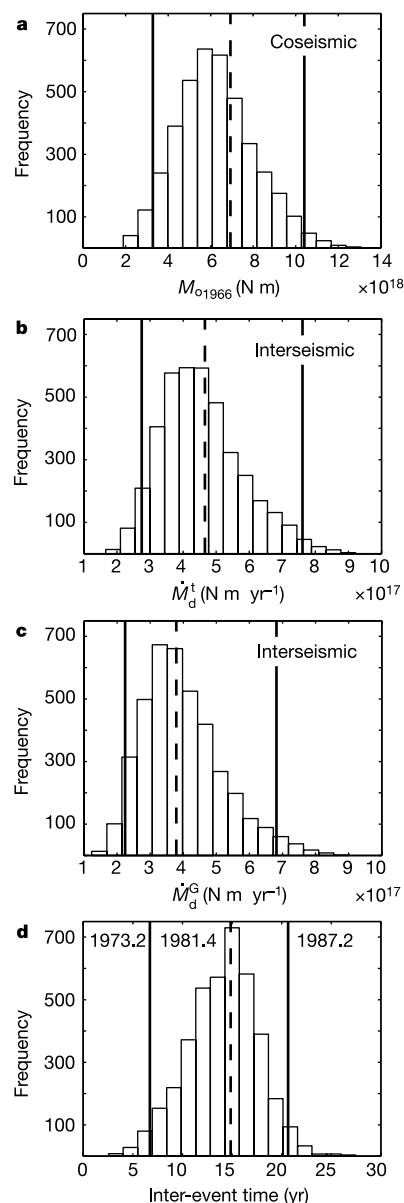


Figure 2 Bootstrap distributions for moment, moment deficit rate, and inter-event time. 95% confidence bounds are indicated by the solid lines. **a**, Coseismic M_o , lower bound is 3.3×10^{18} N m and upper bound is 1.0×10^{19} N m. **b**, Interseismic $\dot{M}_d^I$ found using trilateration data, 2.7×10^{17} N m yr⁻¹ to 7.6×10^{17} N m yr⁻¹. **c**, Interseismic $\dot{M}_d^G$ found using GPS data, 2.2×10^{17} N m yr⁻¹ to 6.8×10^{17} N m yr⁻¹. **d**, Predicted inter-event time, 6.7–20.7 yr. Dashed lines indicate best-fitting value using the original (not resampled) data set.

the fault is assumed to slip aseismically at the long-term rate. There is a well-known trade-off between the estimated transition depth and deep slip-rate; the greater the transition depth the higher is the deep slip-rate^{12,13}. Another complication is that inversion of geodetic data for fault slip is not unique. Regularizing the inversion by, for example, imposing spatial smoothing as in Fig. 1b, produces results contingent on this assumption.

Using data up to 1984, Segall and Harris⁸ found inter-event times from 5 to 29 years. Their analysis was restricted to regularized inversions, and they did not develop a probability distribution of recurrence times. We have analysed nearly twice as much data consisting of (1) coseismic line-length changes, (2) interseismic rates of line-length change (1966–1991), and (3) GPS velocities (1991–1998). We followed Johnson *et al.*¹⁶ and employed constrained inversion to determine M_o and $\dot{M}_d$ independently of assumptions such as smoothness. Bootstrap methods¹⁷ were used to incorporate uncertainty in the transition depth and deep slip-rate and to produce distributions of M_o and of $\dot{M}_d$ from which we determined rigorous bounds on inter-event time.

The range of coseismic M_o at 95% confidence, 3.3×10^{18} N m to 1.0×10^{19} N m (moment magnitude, M_w , of 6.3–6.6), estimated from the bootstrap resamples (Fig. 2a), is consistent with other estimates based on geodetic data^{8,18}, but larger than that found using surface waves¹⁹ (0.9×10^{18} to 2.1×10^{18} N m). The trilateration data spanning the earthquake were collected fairly close to the

fault, allowing large slip at depth without causing large misfit to the data (Fig. 3a). Moreover, considerable afterslip was observed following this event²⁰. Because we are interested in the time needed to accumulate the total moment associated with this earthquake, it is irrelevant whether the slip occurred during fast rupture or post-seismic transient. The $\dot{M}_d$ estimated from the trilateration and GPS data sets individually are comparable, with slightly greater values found using the trilateration data (Fig. 2b, c). The resulting bounds on inter-event time (Fig. 2d) range from 7 to 21 years at 95% confidence. In other words, the strain released in 1966 recovered some time between 1973 and 1987. The expected earthquake has yet to occur.

There are several possible explanations for the time-predictable model's inability to predict the $M \approx 6$ Parkfield earthquake. The model assumes that an earthquake occurs when the fault reaches a critical stress. However, local variations in pore fluid pressure could modulate the stress required for earthquake nucleation. Moreover, formulations for fault failure such as rate and state friction do not embody a characteristic failure stress. Rather, acceleration toward unstable slip depends on the elastic loading system, frictional properties of the fault, and past slip history^{21,22}. Stress perturbations may lead to evolution of the fault 'state', thus affecting the time to the next earthquake²³. For example, Simpson *et al.*²⁴ showed that static stress changes due to the Coalinga earthquake of 1983 may have delayed the next $M \approx 6$ Parkfield earthquake by 1 or 2 years. Including rate and state effects, this delay could extend into the mid-1990s but cannot explain the now 36-year-long quiescence²⁵. Simple stress-change calculations suggest that three $M \geq 4.3$ earthquakes that occurred near the 1966 hypocentre between 1992 and 1994 (Fig. 1b) increased stress at the 1966 hypocentre, countering the effects of the Coalinga quake. There is also evidence for a change in strain rate in the early to mid-1990s (ref. 26).

There may also be a much longer scale decrease in stressing rate following the 1857 Fort Tejon earthquake, which has led to the suggestion that the recurrence times of Parkfield earthquakes should increase with time²⁷. That the recent GPS data predict a deep slip-rate consistent with the geologic rate¹², however, suggests that transient effects due to the 1857 event are currently small. Furthermore, in this study we used actual measurements of strain accumulation post-1966, and our conclusions are independent of longer-term variations in loading. The consistency between GPS and geologically derived deep slip-rates also suggests that any inelastic component of the measured strain is small.

The time-predictable model as conceived by Shimazaki and Nakata¹, and as applied in hazard assessment, assumes repeated rupture of the same fault segment. The 1934 Parkfield earthquake

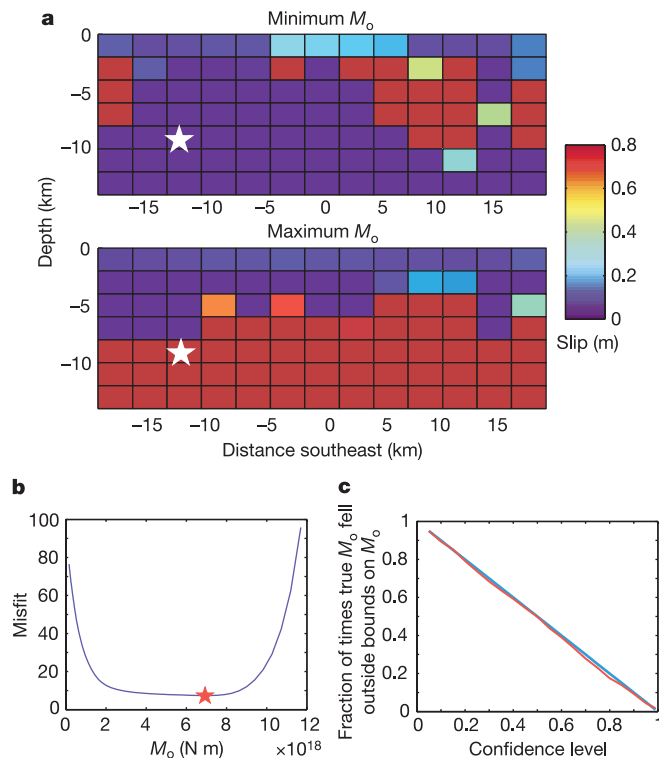


Figure 3 Constrained inversion and bootstrap procedure. **a**, Slip distributions from constrained inversion corresponding to 95% confidence limits on M_o for transition depth of 14 km. The fault plane is same as the gridded plane in Fig. 1b; the star is 1966 hypocentre. For the minimum M_o (3.2×10^{18} N m), the inversion results have most slip at the southeastern end of the fault where the geodetic measurements were concentrated. For the maximum M_o (7.9×10^{18} N m), the results have considerable slip at depth and at the northwestern end where data resolution was poor. **b**, Misfit ($\mathbf{r}^T \Sigma^{-1} \mathbf{r}$) as a function of coseismic M_o . $\mathbf{r}$ is the difference between observed and predicted data, and Σ is the data covariance. **c**, Test of the constrained inversion/bootstrap method for bounding M_o using synthetic data; blue line is predicted behaviour, red line is observed behaviour.

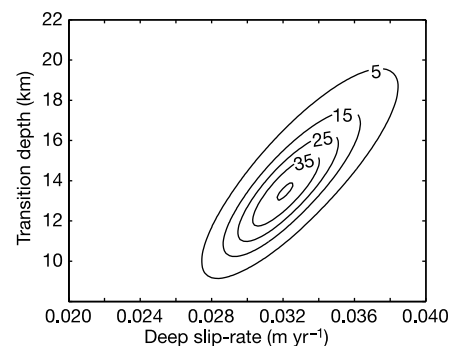


Figure 4 Approximate joint probability density function (indicated on contours) for transition depth and deep slip-rate. The PDF is proportional to e^{-CVSS} (see text) and normalized so that the volume under the surface is one; we note the strong positive correlation between slip-rate and transition depth.

probably ruptured a shorter segment than that in 1966 (ref. 18). However, repeating our test using only estimates of 1966 strain release and subsequent accumulation on the inferred 1934 rupture segment results in nearly identical predicted recurrence times.

Although the time-predictable model does not require successive earthquakes to be of constant size, hazard assessments generally concern the repeat of large earthquakes. The $M \geq 4.3$ earthquakes between 1992 and 1994 were not considered sufficient to fulfill the Parkfield prediction, nor did they rupture the same area as the 1966 quake. These events may indicate that the stress released in 1966 had reaccumulated near the hypocentre by the early 1990s. However, the local stress may depend on details of the creep distribution, so these events do not constrain the stress averaged over the Parkfield segment. The $M = 4.6$ earthquake in 1993 was similar to a foreshock of the 1934 $M \approx 6$ earthquake²⁸. That no $M \approx 6$ event occurred in the early 1990s suggests that conditions for unstable slip may be met locally on the fault, but the earthquake's ultimate size depends on detailed stress distribution and complex rupture dynamics.

Another recurrence model, the slip-predictable model¹, posits that the slip in the next earthquake equals the accumulated slip deficit. This model, which incorporates assumptions similar to those of the time-predictable model (for example, representing earthquakes by only moment and time), cannot be tested until the next $M > 6$ Parkfield quake. However given the current slip deficit rate, were an earthquake to occur on the 1966 rupture plane in 2002 the slip predictable model would require a moment of 9.2×10^{18} N m to 2.6×10^{19} N m ($M_w = 6.6$ – 6.9), considerably larger than the 1966 earthquake.

That the time-predictable model does not accurately predict the anticipated $M \approx 6$ earthquake at Parkfield indicates that there are crucial aspects of earthquake physics, even in a fairly simple and well-understood setting, that this model does not capture. It would be useful to apply the methods developed here to other faults, although to date few have the required history of geodetic observations. Newly developed GPS networks will provide the measurements necessary to determine if Parkfield is representative of the majority of active fault segments. □

Methods

Constrained inversion

Constrained inversion involves finding the slip (or slip deficit rate) distribution that best fits the data, subject to the constraint that the seismic moment (or moment deficit rate) equals a specified value. The model fault (Fig. 1) was based on the mapped fault trace and seismicity¹². Measured surface offsets from the 1966 event were used to bound the estimated coseismic slip in subfaults that break the earth's surface²⁰. Shallow creep rates inferred from near-fault monitoring instruments were used to bound the slip-rates on these subfaults in the inversions for interseismic slip deficit rate¹³. For the interseismic period, the contribution to line-length rate of change or GPS velocity from slip on the creeping section of the San Andreas and from slip below the transition depth were incorporated with uniformly slipping dislocations¹² (note that these dislocations are larger than shown in Fig. 1b). We constrained slip to be positive (right lateral), but made no assumption of smoothness (for example, see Fig. 3a). The upper bound for coseismic slip was 0.8 m, which exceeds the maximum slip estimated in previous studies^{8,18}. The maximum allowable slip deficit rate was the deep slip-rate. We swept through a range of moments and found a best-fitting solution for each one (Fig. 3b). The M_0 with the lowest misfit (star in Fig. 3b) is optimal in the sense that there are no slip distributions resulting in different moments that fit the data better.

The bootstrap

The range of moments that fit the data acceptably well is more important than the best-fitting M_0 . To identify this range, we used the bootstrap, which involves resampling the data a large number of times and repeating the inversions with each resampled data set, resulting in many realizations of the estimated quantity. We repeated the constrained inversions for best-fitting M_0 4,000 times, yielding a distribution of M_0 estimates, and did the same for the interseismic period to infer the distribution of $\dot{M}_d$. We then used these distributions in the calculation for inter-event time.

The distributions can also be used to obtain confidence intervals on M_0 and $\dot{M}_d$. We found this method for inferring the range of M_0 that fit the data to be successful in an empirical test using data predicted from a hypothetical slip distribution (Fig. 3c). We repeated the constrained inversion and bootstrap process 700 times with different random errors added to the synthetic data. The fraction of times the 'true' M_0 fell outside the

inferred range is shown as a function of confidence interval (Fig. 3c, red line). For comparison, the blue line shows the predicted behaviour; for example, at the 95% confidence level the 'true' M_0 should fall outside the range 5% of the time.

Uncertainty in transition depth and deep slip-rate

The estimates of M_0 and $\dot{M}_d$ resulting from the constrained inversion and bootstrap are conditional on an assumed transition depth, and in the case of $\dot{M}_d$, deep slip-rate. Although the geodetic data provide some constraints on these parameters, they are not uniquely resolved. Murray *et al.*¹² estimated interseismic slip-rates for a range of transition depth and deep slip-rate pairs with optimal smoothing determined by cross-validation²⁹. The cross validation sum of squares (CVSS), a measure of misfit, is a nearly quadratic function of transition depth and deep slip-rate. Its minimum is at 14 km and 33 mm yr⁻¹, in keeping with independent geologic and seismic observations. For linear least squares the residual sum of squares (RSS) is quadratic, and the corresponding probability density function (PDF) is proportional to e^{-RSS} . We used the observed distribution of CVSS to generate an approximate PDF proportional to e^{-CVSS} (Fig. 4). Each time the data were resampled in the bootstrap, a new transition depth and deep slip-rate pair was chosen from this empirical distribution.

Calculating inter-event time

The distribution for interevent time (Fig. 2d) was calculated from the ratios of M_0 to $\dot{M}_d$, ensuring that for each pair of M_0 and $\dot{M}_d$ both were found using the same transition depth. Noting that we had 25 years of trilateration measurements without a Parkfield earthquake, we estimate the inter-event time as:

$$\Delta t = \frac{M_{01966}}{\dot{M}_d} \quad (4)$$

if $M_{01966} - (\dot{M}_d \times 25 \text{ yr}) \leq 0$, and

$$\Delta t = 25 \text{ yr} + \frac{M_{01966} - (\dot{M}_d \times 25 \text{ yr})}{\dot{M}_d^G} \quad (5)$$

if $M_{01966} - (\dot{M}_d \times 25 \text{ yr}) > 0$ where the next earthquake is predicted to occur in $1966 + \Delta t$. M_{01966} is the coseismic moment, $\dot{M}_d$ is the moment deficit rate inferred from trilateration data, and $\dot{M}_d^G$ is that inferred from GPS data.

Received 5 March; accepted 15 July 2002; doi:10.1038/nature00984.

- Shimazaki, K. & Nakata, T. Time-predictable recurrence model for large earthquakes. *Geophys. Res. Lett.* **7**, 279–282 (1980).
- Reid, H. F. *The California Earthquake of April 18, 1906* Report of the state earthquake investigation commission Vol. 2 (Carnegie Institute, Washington DC, 1910).
- Working Group on California Earthquake Probabilities *Earthquake probabilities in the San Francisco Bay Region: 2000 to 2030 – A summary of findings* US Geological Survey open-file report 99-517 (USGS, Reston, VA, 1999).
- Working Group on California Earthquake Probabilities Seismic hazards in southern California: probable earthquakes, 1994 to 2024. *Bull. Seismol. Soc. Am.* **85**, 379–439 (1995).
- Cramer, C. H., Petersen, M. D., Tianqing, C., Topozada, T. R. & Reichle, M. A time-dependent probabilistic seismic-hazard model for California. *Bull. Seismol. Soc. Am.* **90**, 1–21 (2000).
- Stirling, M. W. *Proc. 12th World Conference on Earthquake Engineering* (New Zealand Society for Earthquake Engineering, Pergamon, Oxford, 2000).
- Annaka, T. & Yashiro, H. *Risk Analysis* (eds Rubio, J. L., Brebbia, C. A. and Uso, J.-L.) (WIT Press, Boston, 1998).
- Segall, P. & Harris, R. Earthquakes deformation cycle on the San Andreas fault near Parkfield, California. *J. Geophys. Res.* **92**, 10511–10525 (1987).
- Bakun, W. H. & McEvilly, T. V. Recurrence models and Parkfield, California, earthquakes. *J. Geophys. Res.* **89**, 3051–3058 (1984).
- Roeloffs, E. & Langbein, J. The earthquake prediction experiment at Parkfield, California. *Rev. Geophys.* **32**, 315–336 (1994).
- King, N. E., Segall, P. & Prescott, W. Geodetic measurements near Parkfield, California, 1959–1984. *J. Geophys. Res.* **92**, 2747–2766 (1987).
- Murray, J. R., Segall, P., Cervelli, P., Prescott, W. & Svare, J. Inversion of GPS data for spatially variable slip-rate on the San Andreas fault near Parkfield, CA. *Geophys. Res. Lett.* **28**, 359–362 (2001).
- Harris, R. A. & Segall, P. Detection of a locked zone at depth on the Parkfield, California, segment of the San Andreas fault. *J. Geophys. Res.* **92**, 7945–7962 (1987).
- Snay, R. A. Enhancing the geodetic resolution of fault slip by introducing prior information. *Manuscr. Geodetica* **14**, 391–403 (1989).
- Savage, J. C. Equivalent strike-slip earthquake cycles in half-space and lithosphere-asthenosphere Earth models. *J. Geophys. Res.* **95**, 4873–4879 (1990).
- Johnson, H. O., Agnew, D. C. & Hudnut, K. Extremal bounds on earthquake movement from geodetic data: Application to the Landers earthquake. *Bull. Seismol. Soc. Am.* **84**, 660–667 (1994).
- Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap: Monographs on Statistics and Applied Probability* 57 (Chapman and Hall, New York, 1993).
- Segall, P. & Du, Y. How similar were the 1934 and 1966 Parkfield earthquakes? *J. Geophys. Res.* **98**, 4527–4538 (1993).
- Tsai, Y.-B. & Aki, K. Simultaneous determination of the seismic moment and attenuation of seismic surface waves. *Bull. Seismol. Soc. Am.* **59**, 275–287 (1969).
- Smith, S. W. & Wyss, M. Displacement on the San Andreas fault subsequent to the 1966 Parkfield earthquake. *Bull. Seismol. Soc. Am.* **58**, 1955–1973 (1968).
- Rice, J. R. & Ruina, A. L. Stability of steady frictional sliding. *J. Appl. Mech. Trans. ASME* **50**, 343–349 (1983).
- Dieterich, J. & Kilgore, B. Implications of fault constitutive properties for earthquake prediction. *Proc. Natl Acad. Sci. USA* **93**, 3787–3794 (1996).
- Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
- Simpson, R. W., Schulz, S. S., Dietz, L. D. & Burford, R. O. The response of creeping parts of the

- San Andreas Fault to earthquakes on nearby faults: Two examples. *Pure Appl. Geophys.* **126**, 665–685 (1988).
25. Toda, S. & Stein, R. Response of the San Andreas fault to the 1983 Coalinga–Nuñez earthquakes: An application of interaction-based probabilities for Parkfield. *J. Geophys. Res.* **107**, 10.1029/2001JB000172, 2002.
 26. Gao, S., Silver, P. G. & Linde, A. T. Analysis of deformation data at Parkfield, California: Detection of a long-term strain transient. *J. Geophys. Res.* **105**, 2955–2967 (2000).
 27. Ben-Zion, Y., Rice, J. R. & Dmowska, R. Interaction of the San Andreas fault creeping segment with adjacent great rupture zones and earthquake recurrence at Parkfield. *J. Geophys. Res.* **98**, 2135–2144 (1993).
 28. Cole, A. T. & Ellsworth, W. L. Earthquake clustering and the long-term evolution of seismicity near Parkfield, California, 1931–1994. *Seismol. Res. Lett.* **66**, 28 (1995).
 29. Wahba, G. *Spline Models for Observational Data* (Society for Industrial and Applied Mathematics, Philadelphia, 1990).
 30. Ellsworth, W., Waldhauser, F. & Cole, A. A new view of the San Andreas Fault: Implications for earthquake interaction at Parkfield. Extended abstract for the 17th Course of the International School of Geophysics “Fault Interaction by Stress Transfer: New Horizons for Understanding Earthquake Occurrence”, Ettore Majorana Foundation and Centre for Scientific Culture, Erice, Italy, 17–23 June 2000.

Acknowledgements

We thank P. Cervelli, J. Savage, R. Tibshirani, J. Langbein, W. Prescott, J. Svarc and H. Johnson for comments and advice. Funding was provided by Stanford University Graduate Fellowships and a USGS National Earthquake Hazards Reduction Program grant.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M. (e-mail: jrmurray@pangea.stanford.edu).

An unusual oviraptorosaurian dinosaur from China

Xing Xu*, Yen-Nien Cheng†, Xiao-Lin Wang* & Chun-Hsiang Chang†

* Institute of Vertebrate Paleontology & Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China

† Division of Geology, National Museum of Natural Science, Taichung, China

Oviraptorosaurians are an unusual group of theropod dinosaurs, with highly specialized skulls^{1–6}. Here we report a new oviraptorosaurian, *Incisivosaurus gauthieri*, gen. et sp. nov., from the lowest part of the Lower Cretaceous Yixian Formation of China. This oviraptorosaurian displays a number of characters closer to more typical theropods, such as a low skull and toothed jaws, thus greatly reducing the morphological gap between oviraptorosaurs and other theropods^{4–11}. *Incisivosaurus* has a pair of premaxillary teeth resembling rodent incisors and small, lanceolate cheek teeth with large wear facets. These dental features were previously unknown among theropods and suggest a herbivorous diet. The new discovery provides a case of convergent evolution and demonstrates that non-avian theropods were much more diverse ecologically than previously suspected.

Theropoda Marsh, 1881

Maniraptora Gauthier, 1986

Oviraptorosauria Barsbold, 1976

Incisivosaurus gauthieri gen. et sp. nov.

Holotype. IVPP V13326 (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing), an almost complete skull and a partial cervical vertebra (Fig. 1).

Etymology. The generic name refers to the presence of incisor-like premaxillary teeth; the specific name is in honour of J. Gauthier for his contributions to theropod systematics.

Locality and horizon. Lujiatun, Shangyuan, Beipiao City, Liaoning,

China; lowest part of Yixian Formation, older than 128 Myr (ref. 12, Hauterivian).

Diagnosis. An oviraptorosaur displaying the following derived characters: a highly heterodont upper dentition (large incisciform first premaxillary tooth, much smaller, subconical second to fourth premaxillary teeth, and very small lanceolate maxillary teeth); large high-angled wear facets on the mesial margins of the teeth; longitudinal crest on the ventral surface of the basisphenoid; contact between the accessory ventral flanges of the pterygoids; subsidiary ectopterygoid fenestra present; triradiate palatine with very short maxillary process.

Incisivosaurus is a small oviraptorosaur, with a basal skull length of approximately 100 mm. The skull is relatively low, and the length of the snout constitutes 48% of the basal skull length (Fig. 1a). The large, subcircular external naris is positioned high on the snout. A prominent subnasal foramen is visible in lateral view. The oval antorbital fenestra is dorsoventrally tall and anteroposteriorly narrow and the triangular maxillary fenestra is located in the anterior corner of the antorbital fossa. The large orbit is approximately 133% of the antorbital fossa length. The large infratemporal fenestra is subtriangular in outline, and the supratemporal fenestra is long anteroposteriorly. In lateral view, the palate is mostly obscured by the maxilla and jugal; in ventral view, the ventral margins of the premaxillae and maxillae are straight (Fig. 1c), rather than sinuous as in oviraptorids.

The large premaxilla has a shallower main body than that of other oviraptorosaurs^{1,2,5,6}. The maxilla forms the anterior border of the antorbital fossa, which contains a pneumatic fossa dorsally and a small opening ventrally between the large antorbital fenestra and the smaller maxillary fenestra. The posteriormost portion of the ventral margin of the maxilla is inset medially, a feature also seen in the troodontid *Sinovenator*¹³. The nasal is pneumatized as in oviraptorids¹, with a large pneumatic fossa immediately posterior to the external naris (Fig. 1a). Unlike other oviraptorosaurs, therizinosaurids, and basal birds^{6,14}, the nasal is longer than the frontal. The parietal is transversely narrow, and is much shorter than the frontal. The lacrimal is pneumatized, and ‘T’-shaped as in dromaeosaurs and troodontids. The postorbital is ‘T’-shaped, unlike the condition in other oviraptorosaurs. The quadrate is inclined posterodorsally, and has a pneumatic fossa on the caudal surface at midshaft length (Fig. 1b), as in troodontids¹³. The quadratojugal closely abuts the lateral surface of the quadrate, and is positioned relatively dorsally. The jugal is deep and straplike and is situated below the orbit.

The crescentic occiput has long, pendant paroccipital processes (Fig. 1b), as in oviraptorids^{1,15} and a prominent supraoccipital crest. The occipital condyle is smaller than the foramen magnum, which is higher than it is wide. The shallow basisphenoid recess is divided by a longitudinal crest (Fig. 1c). The anteriorly oriented basiptyergoid processes are reduced.

The pterygoid has an accessory ventral flange that contacts its fellow on the midline (Fig. 1c). The ‘C’-shaped ectopterygoid, with its hooked jugal process, extends vertically. It has an extensive contact with the jugal and barely touches the lacrimal. A large palatine fenestra is present posterior to the palatine, and a second much smaller opening (subsidiary ectopterygoid fenestra) is present between the ectopterygoid and the ventral flange of the pterygoid. The triradiate palatine has a large pterygoid process, a moderately large vomeral process, and a very short maxillary process. The internal naris, usually bordered by the palatine, vomer and maxilla in other non-avian theropods, is absent here; instead, it might be confluent with the subsidiary palatine fenestra that is bordered by the vomer and the vomeral process of the palatine.

The mandible is shallow, with a large, anteroposteriorly long mandibular fenestra (Fig. 1d), as in Caenagnathidae^{1–4,16}. The mandibular symphysis is fused, with a small symphyseal shelf. The anterior end of the mandible is toothless and forms a small beak. In side view, the dentary has a slightly convex dorsal margin

and a concave ventral margin. The posterodorsal process of the dentary is short, in contrast to the condition seen in other oviraptorosaurs. However, the posteroventral process of the dentary is long and slender, and reaches the posterior limit of the mandibular fenestra, as in other oviraptorosaurs. The coronoid process is weakly developed and a straplike coronoid appears to be present (Fig. 1e). The surangular extends significantly anteriorly, as in other oviraptorosaurs, but its posterior end is missing. The angular is large, and makes the largest contribution to the posterior part of the mandible. The articular and prearticular are not preserved. The splenial is long and straplike, and its anterior end is expanded dorsoventrally.

The upper dentition is markedly heterodont (Fig. 1a, c). There are four premaxillary and nine maxillary teeth. The greatly enlarged, anteroposteriorly compressed, first premaxillary tooth projects anteriorly and has a large, slightly concave wear facet on the lingual surface that has a sharp enamel cutting edge. *Caudipteryx* also has a pair of long premaxillary teeth, but they are slender and curved^{17,18}. The other three premaxillary teeth of *Incisivosaurus* are much smaller, slender, and are subconical. The maxillary teeth are very small relative to the first premaxillary tooth, are lanceolate (as in therizinosauroids), and lack serrations or denticles. Most teeth have prominent, obliquely inclined wear facets on their mesial margins. There are eight or nine dentary teeth, which are closely packed, and which are similar to the maxillary teeth in size and shape.

The oviraptorosaurian affinities of *Incisivosaurus* are established by the following derived features^{1–6}: skull with short preorbital region; large premaxillary main body; dorsally positioned external naris; pendant paroccipital process; vertically oriented ectopterygoid; fused dentary symphysis; long and shallow posteroventral process of the dentary; strap-like splenial; large mandibular fenestra;

tra; and long retroarticular process. Unlike the short, deep, toothless skulls of most other oviraptorosaurs^{1–6}, *Incisivosaurus* has a relatively low skull and toothed jaws. The shallow mandible of *Incisivosaurus* lacks some of the specializations seen in Oviraptoridae or more inclusive clades within Oviraptorosauria^{1–6}. *Incisivosaurus* displays an intermediate cranial morphology between the typical coelurosaurs and the unusual oviraptorids, and shortens the morphological distance between the two groups. Cladistic analysis places *Incisivosaurus* as the most basal oviraptorosaur (Fig. 2). Collected from sediments deposited more than 128 Myr ago¹², *Incisivosaurus* represents the oldest definitive oviraptorosaur found to date. This is consistent with its basal position in Oviraptorosauria.

Although most work suggests that oviraptorosaurs have a relatively distant relationship to avians^{3,13–15,19}, some recent work has suggested that oviraptorosaurs are close relatives of birds or are even secondarily flightless birds^{6,11,20}. Many characters (including, toothless jaws, short nasals, long parietals, quadrate with a lateral cotyle for the quadratojugal, rodlike jugal bar, long maxillary process of the palatine, absence of a subsidiary palatine fenestra, ectopterygoid that articulates primarily with the lacrimal and maxilla laterally, and absence of a jugal hook on the ectopterygoid) have been used to support the avialan status of oviraptorosaurians^{6,11,20}. We note that the basal oviraptorosaur *Incisivosaurus* lacks these birdlike features, which suggests that they have evolved independently in avialians and derived oviraptorosaurs. Within basal maniraptoran lineages, such as Oviraptorosauria and, possibly, Alvarezsauridae, the derived and stratigraphically younger members are more birdlike in a number of features than the basal taxa; in more derived maniraptoran lineages, including Dromaeosauridae and Troodontidae, the later, derived members become less birdlike by secondary

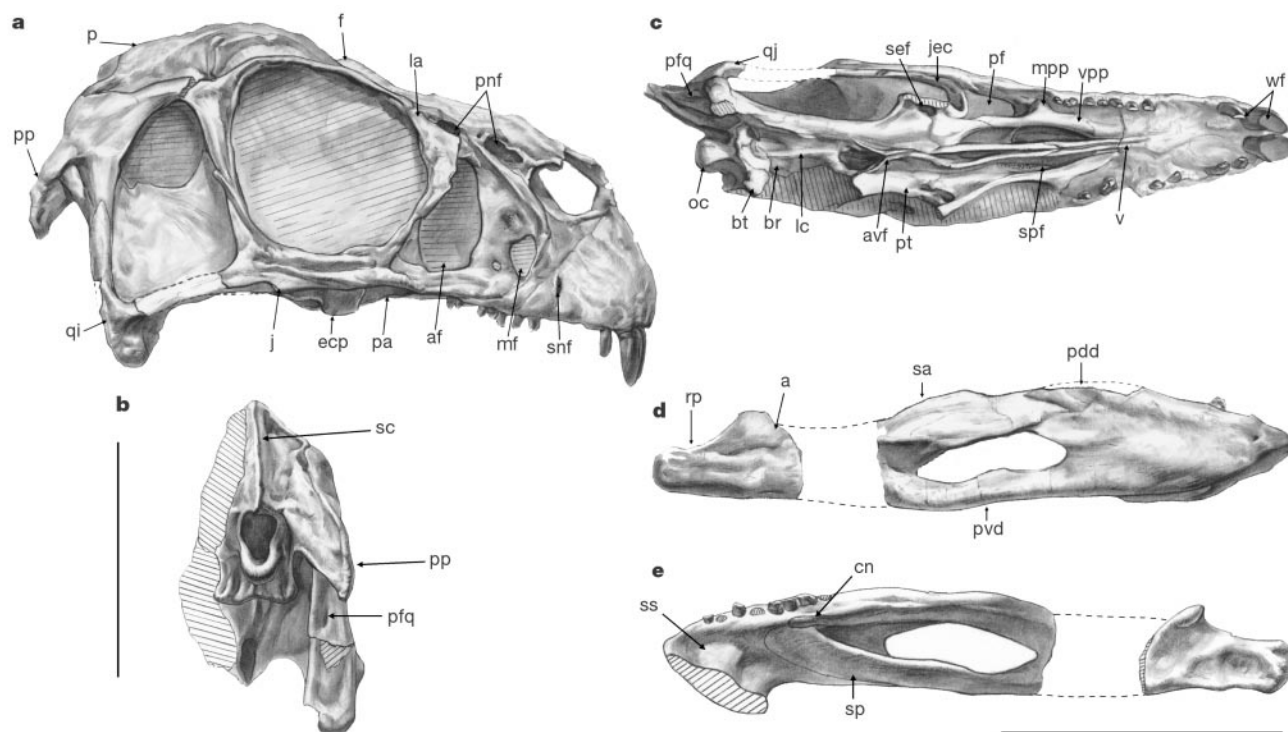


Figure 1 Holotype skull of the oviraptorosaurian *Incisivosaurus gauthieri* (VPP V13326). **a–e**, Lateral (**a**), occipital (**b**), and ventral (**c**) views, and mandible in lateral (**d**) and medial (**e**) views. Scale bar, 4 cm. Abbreviations: a, angular; af, antorbital fenestra; avf, accessory ventral flange; br, basisphenoid recess; bt, basal tubera; cn, coronoid; ecp, ectopterygoid; f, frontal; j, jugal; jec, jugal process of ectopterygoid; la, lacrimal; lc, longitudinal crest; mf, maxillary fenestra; mpp, maxillary process of palatine; oc, occipital condyle; p, parietal;

pa, palatine; pdd, posterodorsal process of dentary; pf, palatine fenestra; pfq, pneumatic fossa on quadrate; pnf, pneumatic fossa; pp, paroccipital; pt, pterygoid; pvd, posterodorsal process of dentary; qj, quadratojugal; rp, retroarticular process; sa, surangular; sc, supraoccipital crest; sef, subsidiary ectopterygoid fenestra; snf, subnasal foramen; sp, splenial; spf, subsidiary palatine fenestra; ss, symphyseal shelf; v, vomer; vpp, vomeral process of palatine; wf, wear facet.

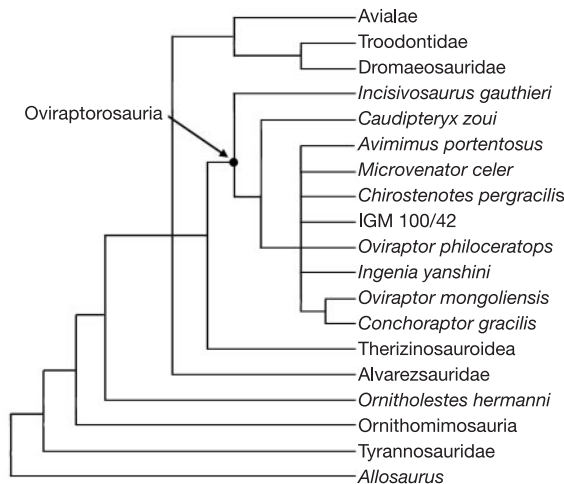


Figure 2 A simplified cladogram representing a strict consensus of 210 trees showing the phylogenetic position of *Incisivosaurus gauthieri*. For detailed information, see Supplementary Information.

reversal of many primitive maniraptoran features. This leads to conflicting results in the reconstruction of maniraptoran phylogeny, and reinforces the importance of including basal members of each group when attempting to reconstruct the phylogeny.

The most interesting features of *Incisivosaurus* are in its dentition. The paired first premaxillary teeth are very similar to the incisors found in a few specialized mammalian lineages, such as rodents, multituberculates and some primatomorphans, which use them for gnawing²¹. The peglike premaxillary teeth are comparable to the dentition of some herbivorous sauripod dinosaurs²²; the lanceolate cheek teeth are similar to those of therizinosaurids, providing further evidence for the close relationship between Oviraptorosauria and Therizinosauridae^{3,13,15}. Although some theropods show a relatively high degree of heterodonty^{13,23–25}, the dental differentiation never reaches the degree seen in *Incisivosaurus* in terms of differences in tooth shape and size. The large wear facets on the teeth indicate tooth-to-tooth occlusion in *Incisivosaurus*. *Incisivosaurus* represents the first theropod displaying distinct dental adaptations for an herbivorous diet, although some other non-avian theropods have been suggested to be herbivorous on the basis of the presence of gastroliths^{17,26} or of some ambiguous morphological evidence^{27–29}.

Received 4 April; accepted 26 June 2002; doi:10.1038/nature00966.

1. Barsbold, R., Maryanska, T. & Osmolska, H. In *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 249–258 (Univ. of California Press, Berkeley, 1990).
2. Barsbold, R. In *The Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 505–509 (Academic Press, San Diego, 1997).
3. Sues, H.-D. On *Chiostenes*, a Late Cretaceous oviraptorosaur (Dinosauria: Theropod) from Western North America. *J. Vert. Paleont.* **17**, 498–716 (1997).
4. Clark, J. M., Norell, M. A. & Barsbold, R. Two new oviraptorids (Theropoda: Oviraptorosauria), Upper Cretaceous Djadokhta Formation, Ukhaa Tolgod, Mongolia. *J. Vert. Paleont.* **21**, 209–213 (2001).
5. Clark, J. M., Norell, M. A. & Rowe, T. Cranial anatomy of *Citipati osmolskae* (Theropoda, Oviraptorosauria), and a reinterpretation of the holotype of *Oviraptor philoceratops*. *Am. Mus. Novit.* **3364**, 1–24 (2002).
6. Maryanska, T., Osmolska, H. & Wolsan, M. Avialan status for Oviraptorosauria. *Acta Palaeontol. Pol.* **47**, 97–116 (2002).
7. Barsbold, R. On a new Late Cretaceous family of small theropods (Oviraptoridae fam. n.) of Mongolia. *Dokl. Akad. Nauk. USSR* **226**, 685–688 (1976).
8. Barsbold, R. Untoothed carnivorous dinosaurs of Mongolia. *Trudy* **15**, 28–39 (1981).
9. Barsbold, R. Carnivorous dinosaurs from the Cretaceous of Mongolia. *Trudy* **19**, 1–120 (1983).
10. Osmolska, H. New light on the skull anatomy and systematic position of *Oviraptor*. *Nature* **262**, 683–684 (1976).
11. Elzanowski, A. A comparison of the jaw skeleton in theropods and birds, with a description of the palate in the Oviraptoridae. *Smithson. Contrib. Paleobiol.* **89**, 311–323 (1999).
12. Swisher, C. C. *et al.* Further support for a Cretaceous age for the feathered-dinosaur beds of Liaoning, China: new ⁴⁰Ar/³⁹Ar dating of the Yixian and Tushengzi Formations. *Chin. Sci. Bull.* **46**, 2009–2013 (2001).

13. Xu, X., Norell, M. A., Wang, X.-L., Makovicky, P.-J. & Wu, X.-C. A basal troodontid from the Early Cretaceous of China. *Nature* **415**, 780–784 (2002).
14. Sereno, P. C. The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999).
15. Makovicky, P. & Sues, H.-D. Anatomy and phylogenetic relationships of the theropod dinosaur *Microvenator celer* from the Lower Cretaceous of Montana. *Am. Mus. Novit.* **3240**, 1–27 (1998).
16. Currie, P. J., Hody, S. J. & Nesson, L. New caenagnathid (Dinosauria: Theropoda) specimens from the Upper Cretaceous of North America and Asia. *Can. J. Earth Sci.* **30**, 2255–2272 (1994).
17. Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
18. Zhou, Z.-H., Wang, X.-L., Zhang, F.-C. & Xu, X. Important features of *Caudipteryx* — evidence from two nearly complete new specimens. *Vert. Palasiat.* **38**, 241–254 (2000).
19. Holtz, T. R. Jr A new phylogeny of the carnivorous dinosaurs. *Gaia* **15**, 5–61 (2000).
20. Paul, G. S. The small predatory dinosaurs of the mid-Mesozoic: the horned theropods of the Morrison and Great Oolite — *Ornitholestes* and *Proceratosaurus* — and the sickle-claw theropods of the Cloverly, Djadokhta and Judith River — *Deinonychus*, *Velociraptor* and *Saurornitholestes*. *Hunteria* **2**, 1–9 (1988).
21. Benton, M. J. *Vertebrate Paleontology* (Chapman & Hall, London, 1997).
22. Upchurch, P. & Barrett, P. M. in *Evolution Of Herbivory In Terrestrial Vertebrates* (ed. Sues, H.-D.) 79–122 (Cambridge Univ. Press, New York, 2000).
23. Sereno, P. C., Forster, C. A., Rogers, R. R. & Monetta, A. M. Primitive dinosaur skeleton from Argentina and the early evolution of Dinosauria. *Nature* **361**, 64–66 (1993).
24. Sampson, S. D., Carrano, M. T. & Forster, C. A. A bizarre predatory dinosaur from the Late Cretaceous of Madagascar. *Nature* **409**, 502–506 (2001).
25. Xu, X., Zhou, Z.-H. & Wang, X.-L. The smallest known non-avian theropod dinosaur. *Nature* **408**, 705–708 (2000).
26. Kobayashi, *et al.* Herbivorous diet in an ornithomimid dinosaur. *Nature* **402**, 480–481 (1999).
27. Russell, D. A. in *Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 729–730 (Academic, San Diego, 1997).
28. Holtz, T. R., Brinkman, D. L. & Chandler, C. L. Denticle morphometrics and a possibly omnivorous feeding habit for the theropod dinosaur *Troodon*. *Gaia* **15**, 159–166 (2000).
29. Barrett, P. M. in *Evolution Of Herbivory In Terrestrial Vertebrates* (ed. Sues, H.-D.) 42–78 (Cambridge Univ. Press, New York, 2000).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank P. Currie, H.-D. Sues, J. Clark, X.-C. Wu, P. Makovicky and P. M. Barrett for reading the manuscript and making valuable comments, M. Norell, C.-K. Li, J.-L. Li, X.-J. Ni and J. Liu for discussions, Z.-H. Zhou for help during the course of the work, H.-J. Wang and Z. Wang for preparing the specimens and R.-S. Li for drawings. We also thank members of the Liaoxi expedition team of the Institute of Vertebrate Paleontology and Paleoanthropology. This work was supported by grants from the National Natural Science Foundation of China, the National Geographic Society, Special Funds for Major State Basic Research Projects of China and the Chinese Academy of Sciences.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.X. (e-mail: xing_xu@sina.com).

Incremental training increases the plasticity of the auditory space map in adult barn owls

Brie A. Linkenhoker & Eric I. Knudsen

Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305-5125, USA

The plasticity in the central nervous system that underlies learning is generally more restricted in adults than in young animals^{1–4}. In one well-studied example, the auditory localization pathway has been shown to be far more limited in its capacity to adjust to abnormal experience in adult than in juvenile barn owls⁵. Plasticity in this pathway has been induced by exposing owls to prismatic spectacles that cause a large, horizontal shift of the visual field. With prisms, juveniles learn new associations between auditory cues, such as interaural time difference (ITD),

and locations in visual space, and acquire new neurophysiological maps of ITD in the optic tectum, whereas adults do neither⁶. Here we show that when the prismatic shift is experienced in small increments, maps of ITD in adults do change adaptively. Once established through incremental training, new ITD maps can be reacquired with a single large prismatic shift. Our results show that there is a substantially greater capacity for plasticity in adults than was previously recognized and highlight a principled strategy for tapping this capacity that could be applied in other areas of the adult central nervous system.

A neurophysiological representation of the associations formed in the central nervous system between auditory localization cues, such as ITD, and locations in space is found in the optic tectum of the barn owl (which is analogous to the mammalian superior colliculus)⁷. In the optic tectum, auditory and visual maps of space are mutually aligned. At a given site in the tectum, neurons respond maximally to auditory or visual stimuli located in the same region of space or receptive field. The mutual alignment of auditory and visual receptive fields signifies that tectal neurons are tuned to the values of auditory localization cues, such as ITD, that correspond correctly to the visual field locations of sound sources⁸.

The auditory–visual associations in the optic tectum are shaped powerfully by early experience^{6,9,10}. For example, when a juvenile owl is exposed to prismatic spectacles (prisms) that displace the visual field horizontally, the tuning of neurons in the optic tectum to ITD (the primary cue for the horizontal position of a sound source) changes over a period of months so that auditory receptive fields realign with the optically displaced visual receptive fields⁶. Because a prominent role of the optic tectum is to control orienting movements¹¹, this plasticity is adaptive because it produces orienting movements in response to auditory stimuli that bring the sound source into the centre of the prismatically displaced visual field of the owl¹².

The capacity of the optic tectum to acquire new representations of auditory cues on the basis of experience is severely restricted in adult owls, as it is in other adult species including mammalian species^{13,14}. Whereas juvenile owls reared with 23° prisms for just a

few months show ITD tuning shifts of 43 μ s (median), adult owls that wear 23° prisms for more than twice as long show ITD tuning shifts of only 4 μ s (median). We verified this severely restricted plasticity in adult owls by exposing seven adult owls aged between 300 d and 4 yr to a single, large prismatic shift. We exposed three owls to 17° prisms and four owls to 23° prisms for at least 2 months, because adjustment to prisms is completed within this time period in juvenile owls.

Figure 1 shows the effect of experiencing 17° prisms for 136 d in one owl. Before prism experience began, ITD tuning in the optic tectum was uniformly sharp and centred on normal values (Fig. 1a). After prism experience, the adaptive (right) flanks of some of the tuning curves had shifted slightly in the adaptive direction (Fig. 1b), but overall there was little difference between the population tuning curves measured before and after this experience (Fig. 1c). The average shift in ITD tuning was $8 \pm 1 \mu$ s (mean $\pm$ s.e.m.). The median shift in ITD tuning among owls exposed to ‘single-step’ 23° prisms was 4 μ s, and the median shift for owls exposed to single-step 17° prisms was 8 μ s. The largest shift in ITD tuning for an owl in either group was 9 μ s.

A second group of adult owls ($n = 5$) was exposed to small incremental increases in prism strength, but otherwise experienced the same conditions as the first group. These owls were aged between 2 and 4 yr and had been through at least two breeding cycles at the start of the experiment. Figure 2a–c shows representative data from one owl exposed to the incremental training model. When prism experience began, ITD tuning was normal. After only 21 d with 6° prisms, ITD tuning had shifted $15 \pm 1 \mu$ s in the adaptive direction (Fig. 2a), a shift that was greater than any shift observed in the owls that experienced single-step prisms for months. The average shift in ITD tuning increased to $20 \pm 1 \mu$ s after an additional 23 d with 11° prisms, and it increased further to $25 \pm 1 \mu$ s after another 42 d with 17° prisms. Because the final shift of 25 μ s was much less than that expected for 17° prisms (43 μ s, see Methods), we decided not to expose the owl to stronger prisms. The magnitude of the final shift was, however, more than twice as large as the largest shift measured in any of the owls experiencing single-step prisms (Fig. 2a, grey zone).

The largest shift in ITD tuning that resulted from incremental training was observed in owl A15 (Fig. 2d–f). ITD tuning in this owl shifted rapidly in response to 6°, 11° and 17° prisms. Because adjustment to 17° prisms was close to the maximum expected shift for 17° prisms (43 μ s), the 17° prisms were replaced with 23° prisms. After 70 d of experience with 23° prisms, the average shift in ITD tuning was $44 \pm 3 \mu$ s relative to normal values (Fig. 2d). The magnitude of this final shift is the same as the median shift that has been observed in juvenile owls exposed to 23° single-step prisms for more than 60 d (ref. 6 and Fig. 2d).

Although the magnitude of the final shifts measured after incremental training varied across individuals (Fig. 3, ‘incremental training’), the increased effectiveness of incremental training was clear. The median shift in ITD tuning after incremental training was 25 μ s, compared with 6 μ s after single-step training, and the smallest shift (19 μ s) was more than twice as large as the largest shift (9 μ s) measured in an owl that experienced single-step prisms (Fig. 3, ‘single step’). Thus, adaptive plasticity was increased by conditions that established a small difference between the existing representation of ITD and the instructed representation.

Owls that had acquired a shifted map of ITD as a result of incremental training could reacquire that shifted map at a later time, in response to a single large step in prism strength. At the end of the incremental training period, prisms were removed and four of the owls experienced normal vision for 29–62 d. During this time, ITD tuning in the optic tectum returned to normal (see Figs 2c, f, shaded curves, and 3, ‘prisms removed’). Owls were then re-exposed to the strongest prisms that they had experienced during their incremental training (23°, owl A15; 17°, owls A11, A17, A18). After

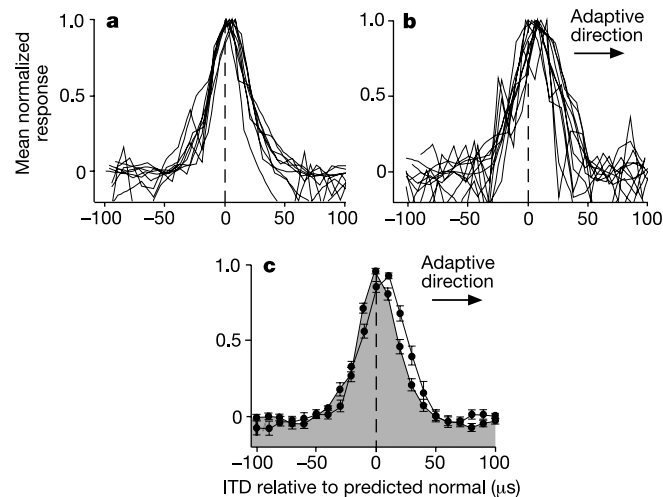


Figure 1 Effect of experience with a single large (17°) step in prism strength on ITD tuning in the optic tectum in one representative owl. **a**, ITD tuning curves from nine sites before prism experience. Average best ITD, $2 \pm 1 \mu$ s (mean $\pm$ s.e.m.). **b**, ITD tuning curves from 11 sites after 136 d with prisms. Average best ITD, $8 \pm 1 \mu$ s. Each curve in **a** and **b** represents the averaged response to 10–20 stimulus repetitions, normalized and plotted as a function of ITD relative to each site’s predicted normal best ITD, as derived from the visual receptive field of the site. **c**, Population ITD tuning curves, generated by averaging data in **a** and **b**. Shaded curve represents ITD tuning before prisms. Line curve represents ITD tuning after prisms. Error bars indicate s.e.m.

33–133 d of experience with these prisms, ITD tuning was evaluated (Fig. 3, 'single-step retest'). The median shift in ITD tuning was $25 \mu\text{s}$; the smallest shift was $18 \pm 2 \mu\text{s}$. The shift measured in each owl was similar to that attained by the end of the incremental training period (compare Fig. 2b with c, and e with f; and Fig. 3 'incremental training' with 'single-step retest'), which indicates that the plasticity resulting from incremental training had expanded the capacity of this pathway for plastic change in a lasting manner.

These results show that the adult auditory localization pathway is capable of far greater adaptive plasticity than has been shown previously⁶. To reveal this capacity, the training increments for adults must be smaller than the training increments for juvenile owls. The need in adults for learning in smaller increments indicates that the visual instructive signal that guides ITD map plasticity operates over a smaller range and/or that it is less effective at producing changes in patterns of connectivity in adult owls relative to juveniles. These effects could result from a developmental decrease in the range and/or plasticity of auditory connections in the localization pathway¹⁵.

Although incremental training is highly effective at inducing adaptive plasticity in adult owls, the maximum range of plasticity is still limited. Small adjustments away from normal tuning occur readily. But larger adjustments decrease in both rate and magnitude, even when the disparity between existing and instructed tuning is small (Fig. 2a, d). The existence of a maximum range of adjustment that is defined relative to normal tuning is consistent with observations that much of the neuronal connectivity that supports normal tuning persists during the learning process¹⁶. This connectivity may restrict functional changes if responses to it cannot be suppressed during the acquisition of learned responses. Restriction of functional plasticity may also result from limited capacities to make distant anatomical connections into new regions of the ITD map¹⁵ or to strengthen existing connections that are near the edge of normal anatomical projection fields.

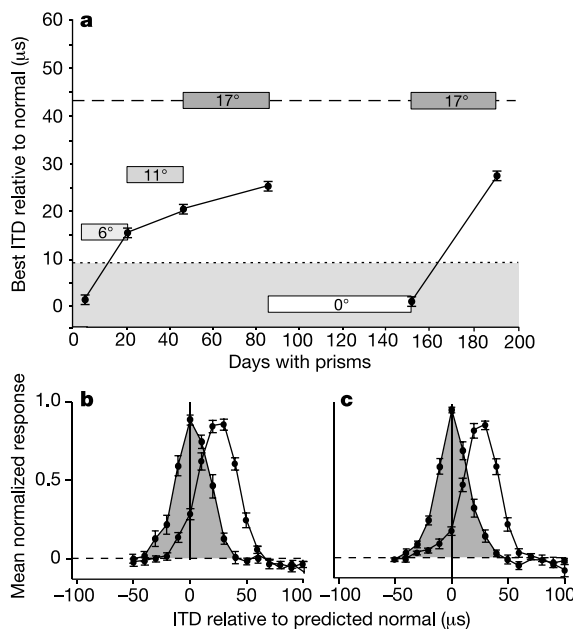


Figure 2 Effects of incremental training on ITD tuning in the optic tectum. **a–c**, Representative effect (owl A11). **a**, Shift in ITD tuning as a function of days of experience. Points represent mean $\pm$ s.e.m. of ITD tuning relative to normal on a given day. ITD shifts corresponding to the prismatic displacements are indicated by the vertical positions of the horizontal bars. Bottom grey zone represents the largest shift in an owl experiencing a single large step in prism strength. Broken line at $43 \mu\text{s}$ represents the median shift in juvenile owls in response to a single large prism step⁶. **b**, Population ITD

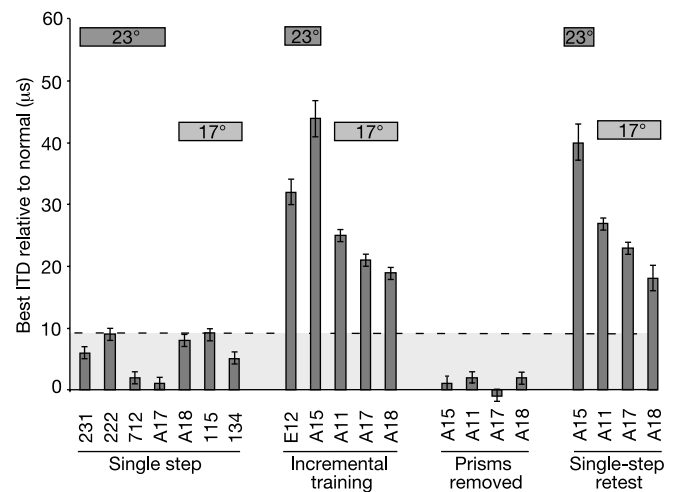
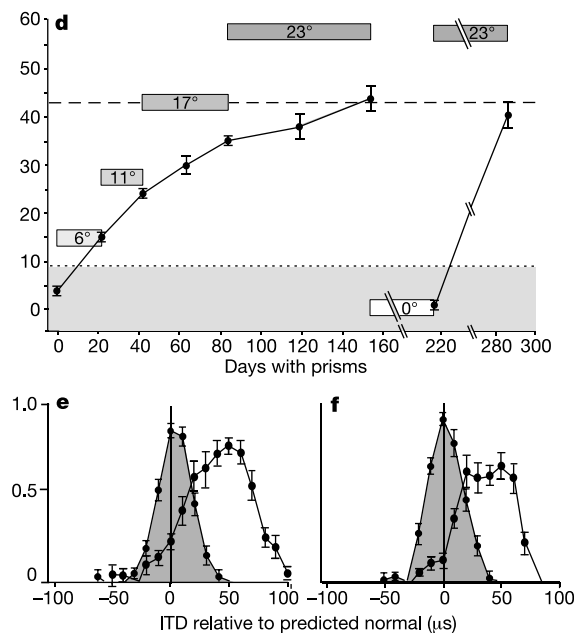


Figure 3 Summary of maximum shifts in ITD tuning. Left to right are shown data from a single large step of the visual field ('single step'), successive small visual field steps ('incremental training'), prism removal after incremental training ('prisms removed'); a single large step after previous incremental training ('single-step retest'). Horizontal bars indicate the strongest prisms experienced by the animal. Bottom grey zone represents the largest ITD shift, averaged across layers, measured in the owl (115) that shifted furthest with a single large step in prism strength. ITD tuning in each owl in the incremental training and retest groups shifted significantly further than in owl 115 ($P < 0.01$, Mann–Whitney, one-tailed test).

The ability of adults to adapt to small incremental changes but not to large ones makes sense in light of the changing demands that are placed on the central nervous system as an animal matures. In young animals, adaptability might be the priority to develop networks that function adaptively despite idiosyncratic or changing



tuning curves (reflecting the average of all ITD tuning curves recorded on a given day) measured before (0 d, shaded) and after (86 d, solid line) incremental training. **c**, Population ITD tuning curves measured before (154 d, shaded) and after (187 d, solid line) experience with a single large (17°) prism step of the visual field after incremental training. **d–f**, Largest effect of incremental training (owl A15), with data plotted as in **a–c**, respectively. **e**, Curves from 0 and 154 d. **f**, Curves from 216 and 286 d. The adaptive direction of ITD shift is to the right in **b**, **c**, **e** and **f**.

properties of the individual or its environment. In older animals, precision and reliability of established, functional networks may be more important.

Adults need to maintain some capacity for adaptive plasticity, however, to permit adjustments to changes in neurological and physical properties that are likely to occur with age. Behavioral studies have shown the power of gradual learning in skill learning, in language acquisition, and in shaping behaviour through operant conditioning^{17–19}. By contrast, our current appreciation of neural plasticity is, for many systems and pathways, based on experiments in which animals are either deprived of experience or are exposed to other large, abrupt changes in experience^{20–23}. Experiments that instead train the nervous system incrementally may uncover substantially greater adaptive neural plasticity in adults as well as in juvenile animals. □

Methods

Animals

Housing in group aviaries, animal preparation, acoustic stimulation and multiunit recording techniques have been described⁶. After prism mounts were cemented to the skull, prism strength could be changed easily by unscrewing a nut that secured the prism spectacle frames to the mount and switching frames. We cleaned prisms twice a week and removed them only for the duration of tectal surveys, which lasted 3–4 h.

Electrophysiology

Multiunit recordings were made in the optic tecta of owls that had recovered from brief halothane anaesthesia. A stainless steel recording chamber was implanted over the optic tectum to permit repeated mapping experiments. Electrophysiological surveys of visual and auditory tuning were restricted to the rostral portion of the optic tectum where visual receptive fields are located between left 10° and right 10° in azimuth, and below 0° in elevation. This portion of visual space is in the centre of the prismatically displaced field and receives extensive stimulation. Auditory tuning in this portion of the tectum shifts reliably in response to prism experience²⁴. We sampled multiunit responses at a minimum interval of 250 µm in the deep tectal layers and 400 µm in the superficial layers, which are distinguished from the deep layers on the basis of a burst discharge pattern²⁵. Electrode penetrations were separated by a minimum interval of 300 µm.

Measurements

Tuning for ITD was measured by delivering computer-generated noise bursts (4 kHz high-pass filtered, 50 ms duration, 0 ms rise/fall time, 1 burst per s, 20–30 dB above unit threshold, optimized interaural level difference) through matched earphones. ITD values, at intervals of 10 µs, were presented 10–20 times in a randomized order. We quantified ITD tuning as the best ITD, which is the centre of the range of ITDs that elicited >50% of the maximum response at each site. The shift in best ITD away from normal was the difference between the measured best ITD and the best ITD expected from the location of the site's visual receptive field. We measured visual receptive fields with dark or light bars projected onto a calibrated tangent screen and recorded the azimuths of their horizontal centres. The dimensions of visual receptive fields in the tectal representation of frontal space are similar across tectal layers, with a width of 10 ± 6° (ref. 25). The relationship between azimuthal angle and ITD has been measured in several owls and is, on average, 2.5 µs per degree of azimuth²⁶. Therefore, to compensate fully for a 23° prismatic shift, ITD tuning should shift by 57 µs. All best ITD values relative to the expected normal values that were recorded on a given day were averaged and are reported as the mean ± s.e.m. throughout this report.

Shifts in ITD tuning away from normal were averaged across the deep and superficial layers of the optic tectum. The capacity for the ITD maps in the superficial and deep layers to shift independently under certain circumstances will be explored elsewhere. We used the Mann–Whitney *U*-test (one-tailed) for all statistical analyses.

Received 30 April; accepted 26 June 2002; doi:10.1038/nature01002.

- Newport, E. L., Bavelier, D. & Neville, H. J. in *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (ed. Doupoux, E.) 481–502 (MIT Press, Cambridge, Massachusetts, 2001).
- Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
- Maurer, D., Lewis, T. L., Brent, H. P. & Levin, A. V. Rapid improvement in the acuity of infants after visual input. *Science* **286**, 108–110 (1999).
- Murphy, E. H., Mize, R. R. & Schechter, P. B. Visual discrimination following infant and adult ablation of cortical areas 17, 18, and 19 in the cat. *Exp. Neurol.* **49**, 386–405 (1975).
- Knudsen, E. I. Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl. *J. Comp. Physiol. A* **185**, 305–321 (1999).
- Brainard, M. S. & Knudsen, E. I. Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci.* **18**, 3929–3942 (1998).
- Knudsen, E. I. Instructed learning in the auditory localization system of the barn owl. *Nature* **417**, 322–328 (2002).
- Brainard, M. S., Knudsen, E. I. & Esterly, S. D. Neural derivation of sound source location: resolution of spatial ambiguities in binaural cues. *J. Acoust. Soc. Am.* **91**, 1015–1027 (1992).
- Gold, J. I. & Knudsen, E. I. Hearing impairment induces frequency-specific adjustments in auditory spatial tuning in the optic tectum of young owls. *J. Neurophysiol.* **82**, 2197–2209 (1999).

- Knudsen, E. I. Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl. *J. Neurosci.* **5**, 3094–3109 (1985).
- Knudsen, E. I., Knudsen, P. F. & Masino, T. Parallel pathways mediating both sound localization and gaze control in the forebrain and midbrain of the barn owl. *J. Neurosci.* **13**, 2837–2852 (1993).
- Knudsen, E. I. & Knudsen, P. F. Vision calibrates sound localization in developing barn owls. *J. Neurosci.* **9**, 3306–3313 (1989).
- King, A. J. A map of auditory space in the mammalian brain: neural computation and development. *Exp. Physiol.* **78**, 559–590 (1993).
- Withington-Wray, D. J., Binns, K. E., Chanjal, S. S., Brickley, S. G. & Keating, M. J. The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental auditory deprivation. *Eur. J. Neurosci.* **2**, 693–703 (1990).
- DeBello, W. M., Feldman, D. E. & Knudsen, E. I. Adaptive axonal remodeling in the midbrain auditory space map. *J. Neurosci.* **21**, 3161–3174 (2001).
- Zheng, W. & Knudsen, E. I. Functional selection of adaptive auditory space map by GABAA-mediated inhibition. *Science* **284**, 962–965 (1999).
- Tallal, P. *et al.* Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* **271**, 81–84 (1996).
- Perone, M., Galizio, M. & Baron, A. in *Human Operant Conditioning and Behavior Modification* (ed. Cullen, G. D. C.) 59–85 (Wiley, New York, 1988).
- Clawson, D. M., Healy, A. F., Ericsson, K. A. & Bourne, L. E. Jr Retention and transfer of morse code reception skill by novices: part-whole training. *J. Exp. Psychol. Appl.* **7**, 129–142 (2001).
- Wiesel, T. N. & Hubel, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).
- Daw, N. W. Critical periods and strabismus: What questions remain? *Optom. Vis. Sci.* **74**, 690–694 (1997).
- Fox, K. A critical period for experience-dependent synaptic plasticity in rat barrel cortex. *J. Neurosci.* **12**, 1826–1838 (1992).
- Mooney, R. Sensitive periods and circuits for learned birdsong. *Curr. Opin. Neurobiol.* **9**, 121–127 (1999).
- Knudsen, E. I. Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science* **279**, 1531–1533 (1998).
- Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
- Knudsen, E. I., Esterly, S. D. & du Lac, S. J. Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls: acoustic basis and behavioural correlates. *J. Neurosci.* **11**, 1727–1747 (1991).

Acknowledgements

We thank Y. Gutfreund and G. Corrado for critical comments on the manuscript. B.A.L. was supported by a National Defense Science and Engineering Graduate Fellowship, a NIH Training Grant, and the Gerald J. Lieberman Fellowship. This research was supported by grants to E.I.K. from The March of Dimes and The National Institute on Deafness and Other Communication Disorders.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.I.K. (e-mail: eknudsen@stanford.edu).

Odorant receptors instruct functional circuitry in the mouse olfactory bulb

Leonardo Belluscio^{*†}, Claudia Lodovichi^{*}, Paul Feinstein[‡], Peter Mombaerts[‡] & Lawrence C. Katz^{*}

^{*} Howard Hughes Medical Institute, Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA
[‡] The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The mammalian olfactory system detects and discriminates thousands of odorants using many different receptors expressed by sensory neurons in the nasal epithelium¹. Axonal projections from these neurons to the main olfactory bulbs form reproducible patterns of glomeruli in two widely separated regions of each bulb, creating two mirror-symmetric maps of odorant receptor projections². To investigate whether odorant receptors organize neural circuitry in the olfactory bulb, we have examined

† Present address: The National Institute of Neurological Disorders and Stroke, National Institutes of Health, 36 Convent Drive, Bethesda, Maryland 20892-4150, USA.

a genetically modified mouse line, $r17 \rightarrow M71$, in which a functionally characterized receptor, $r17^{3,4}$, has been substituted into the M71 receptor locus⁵. Here we show that despite their ectopic location the resulting glomeruli are responsive to known ligands of the $r17$ receptor, attract postsynaptic innervation by mitral/tufted cell dendrites, and endow these cells with responses that are characteristic of the $r17$ receptor. External tufted cells receiving input from $r17 \rightarrow M71$ glomeruli form precise intrabulbar projections that link medial and lateral $r17 \rightarrow M71$ glomeruli anatomically, thus providing a substrate for coordinating iso-functional glomeruli. We conclude that odorant receptor identity in epithelial neurons determines not only glomerular convergence and function, but also functional circuitry in the olfactory bulb.

Axons of olfactory sensory neurons that express the same receptor converge onto two small sets of glomeruli at reproducible locations in each bulb: one on the medial and one on the lateral surface^{6–8}. Receptor-swap experiments have shown that substituting

the expression of one receptor for another produces a positional shift of the corresponding glomeruli, which suggests that odorant receptors at least partly instruct glomerular topography^{8,9}. In the $r17 \rightarrow M71$ mouse line, the rat I7 receptor, which is typically expressed in the ventral epithelium, has been substituted into a dorsally expressed receptor locus, M71, along with green fluorescent protein (GFP) and *tau-lacZ* markers⁵. The resulting $r17 \rightarrow M71$ glomeruli are located ectopically about 600–700 μm anterior and ventral to the endogenous M71 glomeruli⁵; this has allowed us to determine the extent of anatomical and functional information imparted by an odorant receptor when it directs the formation of new glomeruli.

We used optical imaging of intrinsic signals to determine whether the ectopic glomeruli respond to odorant stimuli. Activation patterns elicited with previously defined $r17$ receptor ligands were correlated with the anatomical location of the $r17 \rightarrow M71$ glomeruli, as assessed by epifluorescence visualization of the associated GFP label. In response to octanal, the best studied $r17$ ligand, intrinsic signal maps for all mice ($n = 6$) had distinct, discrete zones of activity that colocalized precisely with the locations of GFP-labelled $r17 \rightarrow M71$ glomeruli (Fig. 1). There was close correspondence between the ligands that activated this focal region and those identified as $r17$ agonists in previous assays of sensory neurons expressing $r17$ (refs 3–5). All intrinsic signal maps of the responses to octanal (0.01–1%) and heptanal (0.01–1%) produced robust activation of the $r17 \rightarrow M71$ glomerulus (Fig. 1b, c).

The ectopic glomeruli were located on the dorsal surface of the bulb, at a considerable distance from the endogenous glomeruli that

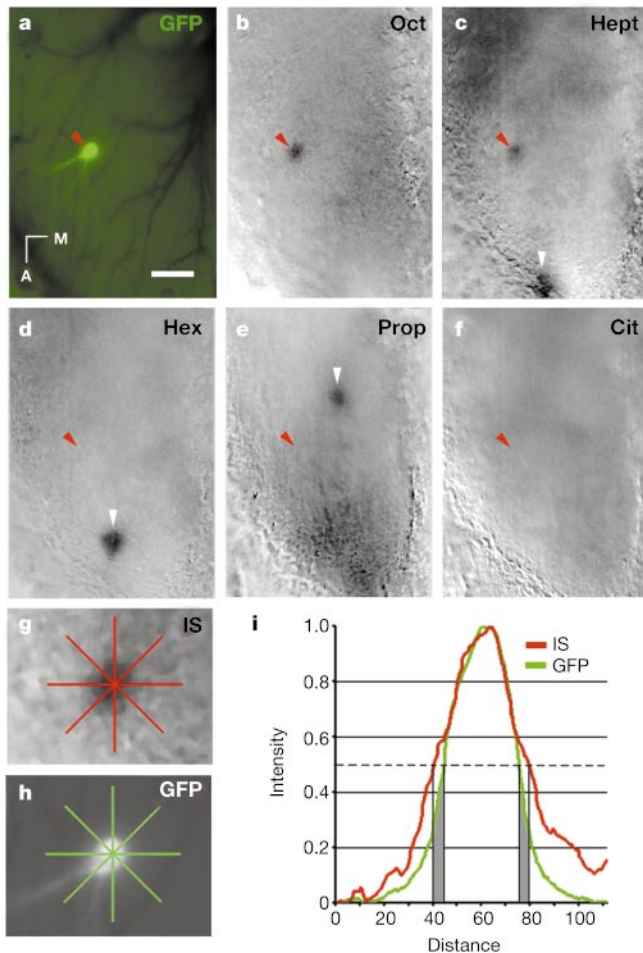


Figure 1 Odorant induced activation of $r17 \rightarrow M71$ glomeruli detected by intrinsic signal imaging. **a–h**, Dorsal view of an adult mouse olfactory bulb imaged *in vivo*. All odorants were presented at 1% concentration. Red arrowheads indicate the location of the $r17 \rightarrow M71$ glomerulus. **a**, Epifluorescent image of an $r17 \rightarrow M71$ GFP-labelled lateral glomerulus (green). **b, c**, Octanal (Oct) and heptanal (Hept) evoke focal $r17 \rightarrow M71$ glomerular activity (dark spots). **d–f**, Hexanal (Hex), propanal (Prop) and citral (Cit) fail to elicit $r17 \rightarrow M71$ activity. White arrowheads in **c–e** mark other regions activated by these odorants¹³. **g, h**, Magnified $r17 \rightarrow M71$ glomerular region from the octanal map (**b**) and GFP image (**a**), each showing four line plots (red and green lines) averaged to determine the intensity profile. **i**, Comparison of normalized intensity profiles of GFP (green) and intrinsic signal (IS, red). The diameters at half-maximum peak intensity show that the intrinsic signal is 22% wider (grey area). Scale bar, 200 μm .

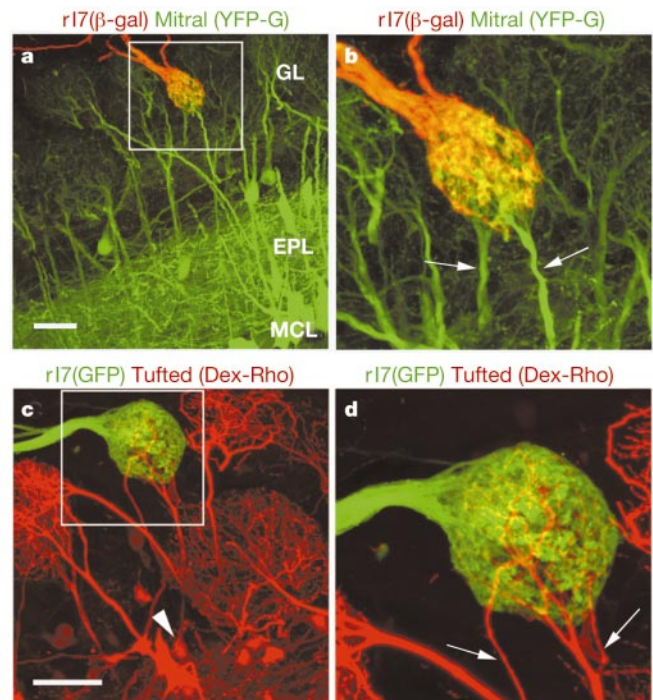


Figure 2 $r17 \rightarrow M71$ glomeruli are innervated by postsynaptic neurons. Shown are confocal microscopy reconstructions of $r17 \rightarrow M71$ glomeruli from coronal sections of the main olfactory bulb. **a**, Immunofluorescent image of an $r17 \rightarrow M71$ glomerulus labelled with *tau-lacZ* and GFP and stained with antibody to β -galactosidase (orange), and YFP-labelled mitral cells (green) show innervation of the $r17 \rightarrow M71$ glomerulus. **b**, Enlargement of the boxed region in **a** shows that the $r17 \rightarrow M71$ glomerulus colocalizes with mitral cell dendrites (arrows). **c**, GFP-labelled $r17 \rightarrow M71$ glomerulus (green) innervated by rhodamine-labelled tufted cells (red). Arrowhead marks the soma of an innervating external tufted cell. **d**, Enlargement of the boxed region in **c** shows colocalization of $r17 \rightarrow M71$ glomerulus and tufted cell dendrites (arrows). Scale bar, 50 μm .

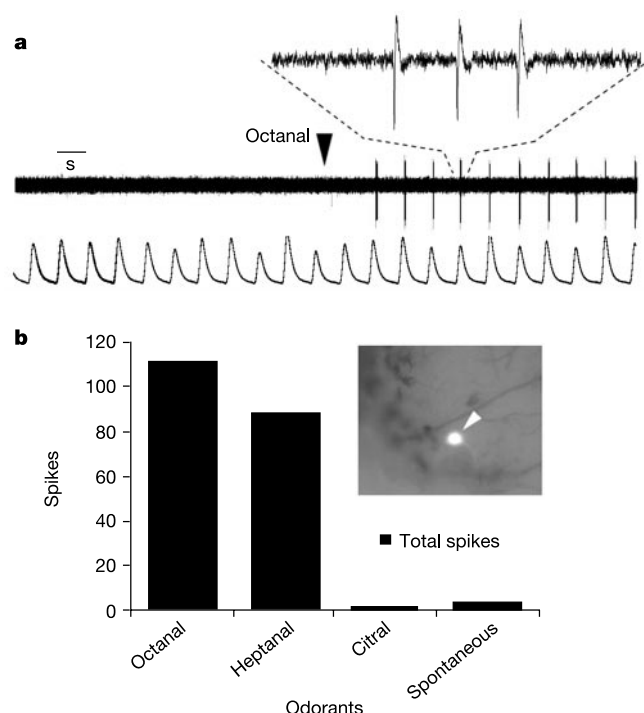


Figure 3 *In vivo* recording from a mitral cell responding to r17 ligands. **a**, In the initial 10 s of recording, the cell has low spontaneous activity. Presentation of octanal (1%, arrowhead) elicits rhythmic bursts (2–4 spikes, inset shows a 60-ms segment) synchronized with breathing rhythm (lower trace). **b**, Total number of spikes for a 30-s combined presentation of odorants shows that robust activity is elicited by octanal and heptanal, but the response to citral (and all other odorants tested, not shown) is comparable to background. Inset shows the recording location, which was located 220 μ m below a GFP-labelled r17 $\rightarrow$ M71 glomerular region (arrowhead).

are activated by octanal and heptanal^{10–14}. By contrast, the structurally related odorant hexanal, which poorly activates the r17 receptor^{3–5}, also failed to induce detectable glomerular activity (Fig. 1d and Table 1). Citral, which activates r17 in one case⁵, but not in others^{3,4}, also failed to elicit a glomerular response. As a positive control, propanal was used to elicit a robust response in each experiment (Fig. 1e); thus, the lack of response to hexanal and citral could not be attributed to a general lack of responsiveness in the imaged bulbs.

The intrinsic signal in response to r17 agonists formed a nearly circular region whose dimensions closely matched those of the GFP-labelled glomerulus. To quantify this relationship, we measured the extent of overlap between the intrinsic signal maps and GFP images. When the average intensity profiles across the r17 region were compared (Fig. 1g, h), the intrinsic signal activated a region 22% larger in diameter than the region outlined by GFP epifluorescence at half-maximum peak intensity (Fig. 1i). As the GFP-labelled sensory afferents occupy only the glomerular neuropil, it is not surprising that the intrinsic signal map, which also incorporates the postsynaptic activity of the surrounding neurons, should occupy a

slightly larger area¹⁴. The close correspondence of these signals shows that the intrinsic signal activity is almost completely confined to a single activated glomerulus.

Having established that olfactory sensory neurons that would normally express the M71 receptor could transmit r17 olfactory information to the bulb, we investigated whether the presence of these ectopic afferents could organize the postsynaptic bulbar circuitry necessary to convey olfactory information to higher regions of the brain. As glomerulus-like structures can form in the absence of postsynaptic olfactory neurons^{15,16}, we wanted to establish whether the dendrites of postsynaptic mitral and tufted cells innervated the r17 $\rightarrow$ M71 glomeruli. We first mated homozygous r17 $\rightarrow$ M71 mice with a Thy1–yellow fluorescent protein (YFP) transgenic line (YFP-G) that expresses YFP in about 30% of mitral cells¹⁷. All of the resulting mice examined showed clear mitral cell innervation of both medial and lateral r17 $\rightarrow$ M71 glomeruli ($n = 5$; Fig. 2). The lateral r17 $\rightarrow$ M71 ectopic glomeruli were typically small in diameter ($35 \pm 2 \mu$ m) as compared with endogenous glomeruli, which average about 75μ m in diameter^{18,19}. Although each r17 $\rightarrow$ M71 glomerulus received input from only 1–2 mitral cells (Fig. 2a, b), this is roughly in proportion to an averaged sized endogenous adult mouse glomerulus receiving input from about 25 mitral cells^{18,19}.

The other main class of projection neurons of the bulb are tufted cells, which, like mitral cells, form a single dendritic tuft that innervates a single glomerulus. When rhodamine-labelled dextran was injected directly into subglomerular regions of the bulb near each r17 $\rightarrow$ M71 glomerulus, all of the dendritic arbors of both mitral and tufted cells were labelled. Confocal reconstruction of these regions clearly showed colocalization of external tufted cell dendrites with r17 $\rightarrow$ M71 glomerular axons, confirming anatomical innervation of ectopic glomeruli by intrabulbar projection neurons (Fig. 2 c, d). Thus, the information conveyed by the exogenous r17 receptor is sufficient to organize the dendritic arbors of the two main types of cell that receive input exclusively from individual glomeruli.

To determine whether the anatomical colocalization described above resulted in the integration of these cells into functional circuits, we made extracellular single-unit recordings from postsynaptic neurons. On the basis of our anatomical findings, we expected that only 3–6 mitral cells would receive input from each ectopic glomerulus. To record from such a small population, we visualized the lateral r17 $\rightarrow$ M71 glomerulus using GFP epifluorescence in live mice and targeted extracellular recordings to their immediate vicinity.

We recorded from 137 mitral cells located near the ectopic glomeruli in 15 mice. Of these, only three neurons (in three different animals) showed responses to octanal (a well-known ligand for the r17 receptor). All three cells had identical response profiles: all had low spontaneous activity and were unresponsive to citral but responded to octanal and heptanal activity with rhythmic, high frequency bursts synchronized to the respiratory cycle (Fig. 3). The responses of these cells were indistinguishable from those of mitral cells that innervate endogenous glomeruli²⁰. These response profiles were also completely consistent with the intrinsic signal response profile of the GFP-labelled glomeruli (Table 1). For example, on the

Table 1 Activity profile of r17 $\rightarrow$ M71 glomerulus

Odorant	Bulb activity	Detected r17 $\rightarrow$ M71 activity*	$\Delta I/I$ at r17 $\rightarrow$ M71–GFP	% of octanal response
Octanal	6/6	6/6	31.0	100
Heptanal	3/3	3/3	29.1	94
Hexanal	3/3	0/3	1.5	5
Citral	0/6	0/6	2.2	7

*All activity maps were aligned with their respective epifluorescent GFP images and initially examined visually for any discernable bulbar activity or r17 $\rightarrow$ M71 activity. The pixel of maximum GFP intensity was located and the corresponding intrinsic map intensity values (I) for a region centred on this pixel were calculated and compared with background to determine $\Delta I/I$ (Methods). The $\Delta I/I$ for each odorant was then compared with the octanal response to obtain a measure of relative intensity.

basis of the aggregate number of spikes, the response to heptanal was about 80% of that elicited by octanal (Fig. 3b and Table 1). Thus, mitral cells acquire the functional identity of the glomerulus that they innervate, which implies that mitral cells can innervate different glomeruli and are not restricted to fixed glomerular targets.

We next investigated the targets of the projection neurons that receive input from the ectopic glomeruli. Because mitral cells project to large ill-defined territories that include several distal regions of the brain^{21,22}, we focused on a specific class of external tufted cells that project between the medial and lateral regions of the olfactory bulb^{21,23,24}. Restricted tracer injections in the lateral half of the bulb revealed restricted intrabulbar projections to the medial

side of the bulb, which suggested that there is a topographical linkage between the two halves of the bulb (refs 23, 24 and Fig. 4a, b). As each hemibulb contains a mirror image map of odorant receptors^{6,7}, we determined whether these projections link isofunctional regions.

Minute tracer injections confined mainly to labelled r17 → M71 glomeruli on the dorsolateral surface ($n = 12$) resulted in a specific projection that terminated in a remarkably restricted region of the granule cell layer on the opposite side of the bulb, centred directly beneath the medial GFP-labelled r17 → M71 glomerulus (Fig. 4c, d). When reconstructed over several adjacent sections, the projections formed a disc with an average diameter of $127 \pm 11 \mu\text{m}$ ($n = 12$). This target region corresponds to the size of the region that contains the complement of mitral cells that innervate an individual glomerulus²⁵. To quantify the relationship between the axonal projections and the r17 → M71 glomerulus, we measured the angle of a line connecting the centroid of each ($n = 12$) and compared this with a line drawn through the glomerulus perpendicular to the pial surface, which revealed a near perpendicular relationship (offset $3 \pm 1^\circ$). Previous studies have shown that in the dorsocaudal region of the bulb, mitral cells project their apical tufts nearly perpendicular to glomeruli residing directly above, consistent with our measurements²⁵. Therefore, it is likely that the antero-grade projection is precisely targeted to the region directly beneath the mitral cells that innervate the corresponding glomerulus in the mirror-image map. In addition, this projection, as in normal animals, is reciprocal between the two mirror-image glomeruli, thereby forming a reciprocal inhibitory circuit (refs 23, 24; and C.L., unpublished data).

Given the regional variability that is typically shown by any identifiable glomeruli^{26,27}, one interpretation of this finding is that expression of the r17 odorant receptor in primary olfactory neurons is sufficient to reorganize the postsynaptic circuitry of the bulb purely through an activity-dependent mechanism such that isofunctional glomeruli remain functionally linked^{21,28}. We consider it more likely, however, that both presynaptic afferents and the axonal projections of postsynaptic tufted cells are drawn to coincident locations in the bulb by common molecular cues that span both the internal and external plexiform layers. Coordinated activity by glomerular pairs may then further refine or maintain the accuracy of this projection. Thus, despite the difference in molecular and cellular environment into which the r17 receptor has been introduced—which cross both species and regional epithelial boundaries⁵—its presence directed the formation of functional glomeruli, recruited postsynaptic dendrites and conveyed r17 function to neurons that were previously unrelated to r17 function. This indicates that postsynaptic neurons may not be obligated to innervate specific predetermined glomeruli, as shown in the *Drosophila* olfactory system²⁹, but can acquire their functional identity by the properties of the sensory axons that converge to form glomeruli near their dendrites. □

Methods

Transgenic mouse lines

We carried out experiments on the r17 → M71 mouse strain in which the mouse M71 odorant receptor coding sequence has been replaced with the coding sequences for the rat I7 odorant receptor, GFP and *tau-lacZ* using homologous recombination⁵. Mice were aged 8–20 weeks. Transgenic Thy1-YFP-G mice were generated using *Thy1* regulatory elements as described¹⁷. Mice expressing the *Thy1*-YFP-G transgene were mated with homozygous I7 → M71 mice to produce I7 → M7 YFP-G mice that expressed the *Thy1*-YFP-G transgene and were heterozygous for the I7 → M71 swap mutation. Mice were maintained in accordance with NIH and institutional animal care guidelines.

Signal imaging and GFP *in vivo* epifluorescence

We recorded intrinsic signals using an Imager 2001 system (Optical Imaging) as described^{10,13}. Epifluorescent images were acquired directly through the Imager 2001 CCD camera using a custom-built GFP imaging system (excitation 488 nm, emission 522 nm).

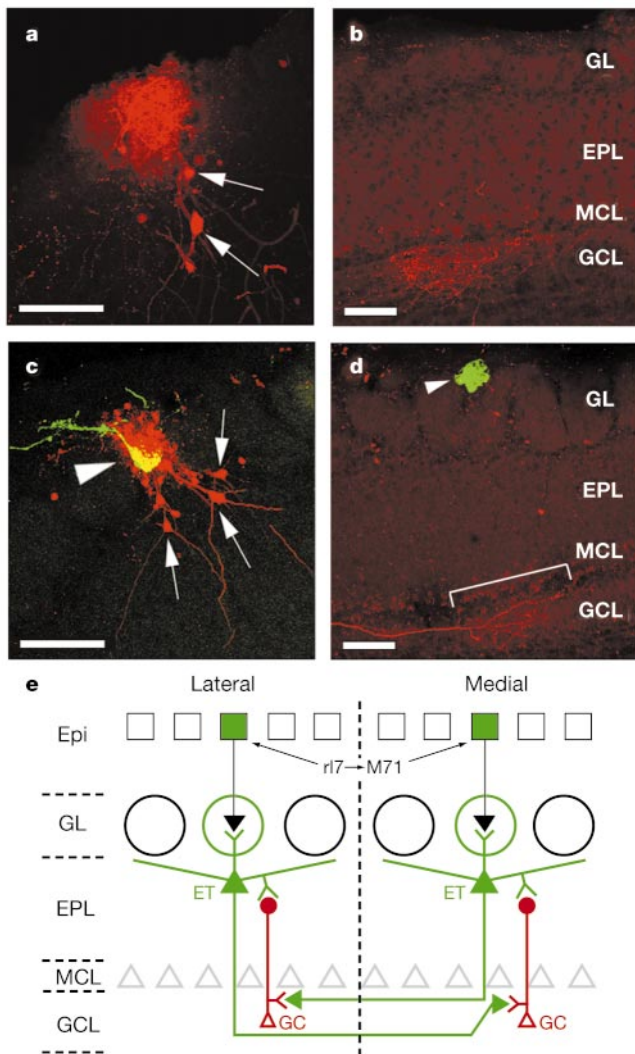


Figure 4 Reciprocal connection between isofunctional glomeruli. **a, b**, Micro-injection of tracer in the dorsolateral bulb (**a**) labels local external tufted cells (arrows), which form a restricted projection in the superficial granule cell layer in the medial bulb (**b**). **c, d**, Tracer injection targeted to a GFP-labelled r17 → M71 glomerulus (**c**, green) on the dorsolateral bulb labels both associated external tufted cells (arrows) and a specific intrabulbar projection (**d**, bracketed region), which is centred beneath the medial r17 → M71 glomerulus. **e**, Diagram of the reciprocal connection that links isofunctional glomeruli. r17 → M71 sensory neurons (green squares) form excitatory synapses (triangle) with external tufted cells (green) in the r17 → M71 glomerulus, which in turn project to the opposite side of the bulb. There, they ramify onto superficial granule cells (red), located directly below the isofunctional r17 → M71 glomerulus, which then form inhibitory synapses (circle) on overlying mitral or tufted cells^{23,24}. Epi, epithelium; GC, granule cell; GCL, granule cell layer; GL, glomerular layer. Scale bars, 50 μm .

Electrophysiology

After carrying out epifluorescent localization of GFP-labelled r17 → M71 glomeruli in anaesthetized mice, we removed the bone over the olfactory bulb and targeted glass electrodes (2–10 MΩ, filled with 2 M sodium acetate) to r17 → M71 glomerular regions. We made 10–15 electrode penetrations in each bulb to locate mitral cells associated with r17 → M71. We isolated 137 cells from 15 mice at the characteristic mitral cell depth of 200–250 μm below the bulbar surface. All isolated cells were tested for response to octanal, heptanal and citral as described^{10,13}. Data were digitally recorded at 30 kHz and analysed using Spike2 software (Cambridge Electronic Design Limited).

Tracer injections

Mice were anaesthetized (isoflurane, 1–3%) and the bone over the r17 → M71 glomerular region was removed. An epifluorescence image was acquired to visualize the GFP-labelled r17 → M71 glomerulus and its precise location was determined with respect to the surface map of blood vessels. Microinjections of rhodamine-labelled dextran, MW3000 (Molecular Probes) were targeted to either the r17 → M71 glomerulus for tracing intrabulbar projections or to a region 150 μm beneath the glomerulus for tracing dendritic innervation. The tracer was iontophoresed (+2 μA, pulsed for 2–5 min) into the olfactory bulb through a micropipette. Mice were killed 8–12 h after injection for histological processing.

Histology and immunohistochemistry

Mice were anaesthetized 8–12 h after tracer injection (Nembutal, 100 mg/kg, i.p.) and perfused transcardially with PBS and 4% paraformaldehyde in 0.1 M phosphate buffer. Olfactory bulbs were removed, post-fixed (2 d) and cryoprotected (30% sucrose for 3 d). Bulbs were then sectioned on a freezing microtome at 40–50 μm, washed (PBS) and mounted (Vectashield, Vector Laboratories). To visualize the r17 glomerulus, the r17 → M71 YFP-G tissue was immunostained as floating sections with primary antibody against β-galactosidase (rabbit polyclonal, Cappel) at 1:1,000 dilution overnight at 4 °C. Cy3-conjugated secondary antibodies against goat (Jackson ImmunoResearch) were used at 1:500 dilution for 2 h at 23 °C.

Confocal imaging

Sections were imaged on a Zeiss Axioskop 2 using a laser scanning confocal microscope (Zeiss LSM 510). Images were obtained with a 40 × (1.3 NA) oil objective for reconstructing individual glomeruli and 25 × (0.8 NA) oil objective for intrabulbar projections. YFP and GFP were excited with the 488-nm line of an argon laser and detected using a 522-nm emission filter. We acquired z-series projections using Zeiss LSM software.

Image analysis

We superimposed GFP images directly on intrinsic signal maps using IPLab (Scanalytics) to observe the colocalization of r17 → M71 glomerulus with odorant activity. Signal intensity (I_s) measurements were determined by averaging a region of 5 × 5 pixels centred over the GFP peak pixel location for each experiment. Background values (I_b) were obtained at a distance of 50 pixels medial from the GFP peak. These values were used to calculate $\Delta I/I$ for each odorant. Intensity line profile measurements were obtained using NIH Image, and pixel values exported to Excel for graph-making and analysis. Measurements of glomerular size and intrabulbar projections were acquired using Zeiss LSM software and all statistical calculations were done in Excel.

Received 29 January; accepted 24 June 2002; doi:10.1038/nature01001.

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. Buck, L. B. The molecular architecture of odor and pheromone sensing in mammals. *Cell* **100**, 611–618 (2000).
3. Zhao, H. *et al.* Functional expression of a mammalian odorant receptor. *Science* **279**, 237–242 (1998).
4. Arandeda, R. C., Kini, A. D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nature Neurosci.* **3**, 1248–1255 (2000).
5. Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22**, 3033–3043 (2002).
6. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
7. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).
8. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
9. Wang, F., Nemes, A., Mendelsohn, M. & Axel, R. Odorant receptors govern the formation of a precise topographic map. *Cell* **93**, 47–60 (1998).
10. Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
11. Johnson, B. A. & Leon, M. Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *J. Comp. Neurol.* **422**, 496–509 (2000).
12. Uchida, N., Takahashi, Y. K., Tanifuji, M. & Mori, K. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nature Neurosci.* **3**, 1035–1043 (2000).
13. Belluscio, L. & Katz, L. C. Symmetry, stereotypy, and topography of odorant representations in mouse olfactory bulbs. *J. Neurosci.* **21**, 2113–2122 (2001).
14. Meister, M. & Bonhoeffer, T. Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.* **21**, 1351–1360 (2001).
15. Graziadei, P. P., Levine, R. R. & Graziadei, G. A. Regeneration of olfactory axons and synapse formation in the forebrain after bulbectomy in neonatal mice. *Proc. Natl Acad. Sci. USA* **75**, 5230–5234 (1978).
16. Bulfone, A. *et al.* An olfactory sensory map develops in the absence of normal projection neurons or GABAergic interneurons. *Neuron* **21**, 1273–1282 (1998).

17. Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
18. LaMantia, A. S. & Purves, D. Development of glomerular pattern visualized in the olfactory bulbs of living mice. *Nature* **341**, 646–649 (1989).
19. Buck, L. B. Information coding in the vertebrate olfactory system. *Annu. Rev. Neurosci.* **19**, 517–544 (1996).
20. Luo, M. & Katz, L. C. Response correlation maps of neurons in the mammalian olfactory bulb. *Neuron* **32**, 1165–1179 (2001).
21. Shipley, M. T. & Ennis, M. Functional organization of olfactory system. *J. Neurobiol.* **30**, 123–176 (1996).
22. Zou, Z., Horowitz, L. F., Montmayeur, J. P., Snapper, S. & Buck, L. B. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **414**, 173–179 (2001).
23. Schoenfeld, T. A., Marchand, J. E. & Macrides, F. Topographic organization of tufted cell axonal projections in the hamster main olfactory bulb: an intrabulbar associational system. *J. Comp. Neurol.* **235**, 503–518 (1985).
24. Liu, W. L. & Shipley, M. T. Intrabulbar associational system in the rat olfactory bulb comprises cholecystokinin-containing tufted cells that synapse onto the dendrites of GABAergic granule cells. *J. Comp. Neurol.* **346**, 541–558 (1994).
25. Buonviso, N., Chaput, M. A. & Scott, J. W. Mitral cell-to-glomerulus connectivity: an HRP study of the orientation of mitral cell apical dendrites. *J. Comp. Neurol.* **307**, 57–64 (1991).
26. Strotmann, J., Conzelmann, S., Beck, A., Feinstein, P., Breer, H. & Peter Mombaerts Local permutations in the glomerular array of the mouse olfactory bulb. *J. Neurosci.* **20**, 6927–6938 (2001).
27. Schaefer, M. L., Finger, T. E. & Restrepo, D. Variability of position of the P2 glomerulus within a map of the mouse olfactory bulb. *J. Comp. Neurol.* **436**, 351–362 (2001).
28. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
29. Jefferis, G. S., Marin, E. C., Stocker, R. F. & Luo, L. Target neuron prespecification in the olfactory map of *Drosophila*. *Nature* **414**, 204–208 (2001).

Acknowledgements

We thank G. Feng for providing the YFP-G mice. L.C.K. is an Investigator in the Howard Hughes Medical Institute, L.B. is a Burroughs Wellcome Fellow in Neuroscience, and grant support to P.M. was from the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.B. (e-mail: belluscl@ninds.nih.gov).

Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier

M. Charles Liberman^{*†}, JIANGANG GAO^{†‡}, DAVID Z. Z. HE[§], XUDONG WU^{‡||}, SHUPING JIA[§] & JIAN ZUO[‡]

^{*} Department of Otology and Laryngology, Harvard Medical School and Eaton-Peabody Laboratory, Massachusetts Eye & Ear Infirmary, Boston, Massachusetts 02114, USA

[‡] Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

[§] Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

^{||} Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA

[†] These authors contributed equally to this work

Hearing sensitivity in mammals is enhanced by more than 40 dB (that is, 100-fold) by mechanical amplification thought to be generated by one class of cochlear sensory cells, the outer hair cells^{1–4}. In addition to the mechano-electrical transduction required for auditory sensation, mammalian outer hair cells also perform electromechanical transduction, whereby transmembrane voltage drives cellular length changes at audio frequencies *in vitro*^{5–7}. This electromotility is thought to arise through voltage-gated conformational changes in a membrane protein^{8,9}, and prestin has been proposed as this molecular motor^{10–12}. Here we show that targeted deletion of prestin in mice results in loss of outer hair cell electromotility *in vitro* and a

40–60 dB loss of cochlear sensitivity *in vivo*, without disruption of mechano-electrical transduction in outer hair cells. In heterozygotes, electromotility is halved and there is a twofold (about 6 dB) increase in cochlear thresholds. These results suggest that prestin is indeed the motor protein, that there is a simple and direct coupling between electromotility and cochlear amplification, and that there is no need to invoke additional active processes to explain cochlear sensitivity in the mammalian ear.

To test the hypotheses that prestin is the molecular motor underlying outer hair cell (OHC) electromotility and that OHC electromotility is the basis of the cochlear amplifier, we created a mutant mouse in which the *prestin* gene was targeted for deletion. Our targeting strategy (Fig. 1a) removed exons 3–7, which encode the 245 amino-terminal amino acids (about one-third of the protein). This strategy removed the ATG start codon, a highly conserved STAT motif, positively charged residues crucial for intracellular anion binding, and the first five of the twelve putative transmembrane domains¹². Genomic Southern analysis (using both external and neomycin (Neo) probes) and polymerase chain reaction (PCR) analysis confirmed the targeted deletion of the *prestin* gene (Fig. 1b–d). Real-time PCR with reverse transcription (RT-PCR) analysis of the deleted coding region showed that the amount of prestin messenger RNA in heterozygous (+/–) cochleae was roughly half that in normal (+/+) cochleae, whereas no transcript was detected in cochleae from prestin null mice (–/–) (Fig. 1e). Immunohistochemistry with antibodies to the N- and carboxy-terminal segments of prestin¹¹ produced no signal in prestin null mice and reduced labelling in OHCs of heterozygous mice relative to that in wild-type mice (Fig. 2d–f; and data not shown).

The mutant mice had no obvious developmental or behavioural abnormalities, except for a small size difference: at 1 month, body weights of sex-matched prestin null ($n = 14$) and heterozygous ($n = 14$) mice were 85% and 95% of wild-type ($n = 16$) littermates,

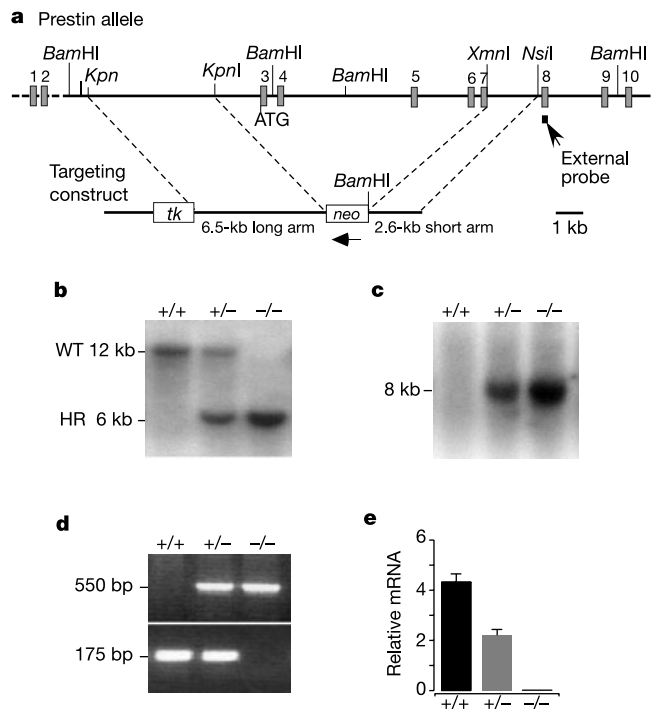


Figure 1 Targeted disruption of the prestin locus. **a**, Strategy for targeted deletion of the prestin gene. *tk*, thymidine kinase gene. **b**, Southern analysis of genomic DNA from wild-type (+/+), heterozygous (+/–) and prestin null (–/–) mice by using an external probe and *Bam*HI digestion. WT, wild type; HR, homologous recombination. **c**, Southern analysis with the Neo probe and *Bam*HI digestion. **d**, PCR genotyping of mutant mice. **e**, Real-time RT-PCR analysis of cochlear prestin mRNA in three genotypes, normalized to a standard sample. Means and standard errors shown (+/+, $n = 12$; +/–, $n = 9$; –/–, $n = 6$).

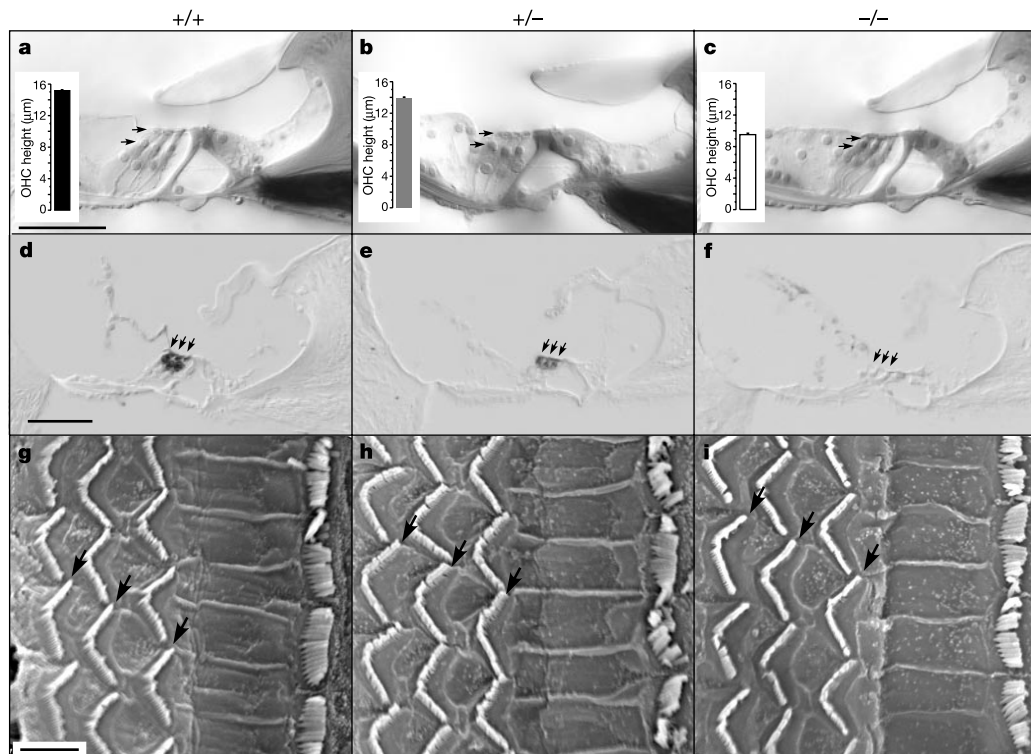


Figure 2 Morphological and immunohistochemical analysis of mutant mice. **a–c**, Light micrographs of the upper basal turn (cochlear frequency ~16 kHz) in cochleae of wild-type, heterozygous and null mice. Arrows at the cuticular plate and nucleus of third-row OHCs illustrate length differences in the three genotypes. Insets show mean lengths ($\pm$ s.e.m.) of over 50 OHCs in the upper basal turn. Scale bar, 50 μ m. **d–f**, Immunostaining for prestin N terminus shows three rows of darkly stained OHCs (arrows) in wild-type cochleae, less intense staining in heterozygous cochleae, and no prestin staining in ears of prestin null mice. Images are from the upper basal turn. Scale bar, 50 μ m. **g–i**, Scanning electron micrographs show normal hair bundles on three rows of OHCs (arrows) from each of the three genotypes. Scale bar, 5 μ m.

Immunostaining for prestin N terminus shows three rows of darkly stained OHCs (arrows) in wild-type cochleae, less intense staining in heterozygous cochleae, and no prestin staining in ears of prestin null mice. Images are from the upper basal turn. Scale bar, 50 μ m. **g–i**, Scanning electron micrographs show normal hair bundles on three rows of OHCs (arrows) from each of the three genotypes. Scale bar, 5 μ m.

respectively. Cochlear morphology (Fig. 2a–c) in prestin null mice at 7–9 weeks of age was normal except for a decrement in OHC lengths in all cochlear turns (insets, Fig. 2a–c) and an almost complete loss of inner hair cells (IHCs) and OHCs in the basal 25% of the cochlear spiral (data not shown). The heterozygous mice had intermediate OHC lengths (inset, Fig. 2b) and showed scattered loss of IHCs and OHCs in the basal 25% of the cochlea. The finding of shortened OHCs is not surprising given that prestin is a principal component of the cell's lateral membranes¹³. The hair cell loss may reflect a developmental role for prestin in ion transport, as suggested by its homology to other transporters¹⁴—hair cell loss was not seen at post-natal day 7 (data not shown). Although prestin is expressed only in OHCs, IHC loss subsequent to primary OHC damage is well documented in the literature of acoustic injury¹⁵.

Further morphological analysis suggested that the hair bundles, which house the machinery for mechano-electrical transduction, were normal in our mutant mice. Scanning electron micrographs

showed normal hair bundle morphology on all rows of OHCs (Fig. 2g–i), and immunofluorescence showed that two principal components of the hair bundle machinery, myosin Ic and myosin VIIA^{16,17}, were present in OHCs of prestin null mice (data not shown).

To determine whether prestin is necessary for normal electro-motility, OHCs were isolated from mice at 5–7 weeks of age. Normal mammalian OHCs show length changes *in vitro* when transmem-

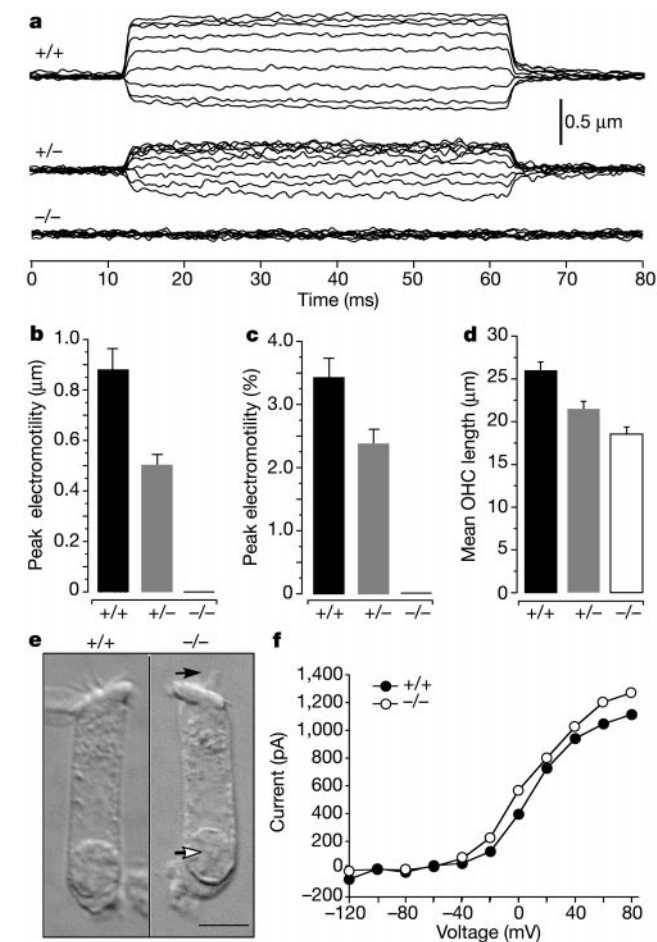


Figure 3 *In vitro* analysis of OHC electromotility in mutant mice. **a**, Length changes of OHCs of wild-type, heterozygous and null mice in response to voltage steps (–120–60 mV in 20-mV steps) in whole-cell, voltage-clamp recordings. Each trace is an average of five stimulus presentations. Cell contraction is plotted upward. **b**, Maximum motility measures for OHCs, expressed as absolute length change. **c**, Maximum motility for OHCs expressed as per cent length change. **d**, Lengths of isolated OHCs of each genotype. Sample sizes for **b–d**, 25 cells from +/+, 21 cells from +/- and 16 cells from -/-. Means and standard errors are shown. **e**, Micrographs of OHCs isolated from the apical turn of wild-type or null cochleae, with nucleus (open arrow) and stereocilia bundle (filled arrow) indicated. Scale bar, 5 μ m. **f**, Current–voltage curves for representative OHCs from wild-type and null mice. The cells were held at –70 mV. *I–V* curves were obtained from measuring steady-state responses after the leak current was subtracted out using the *Pf* – 4 technique. About 70% of series resistance was corrected. The uncompensated voltage error was less than 5 mV.

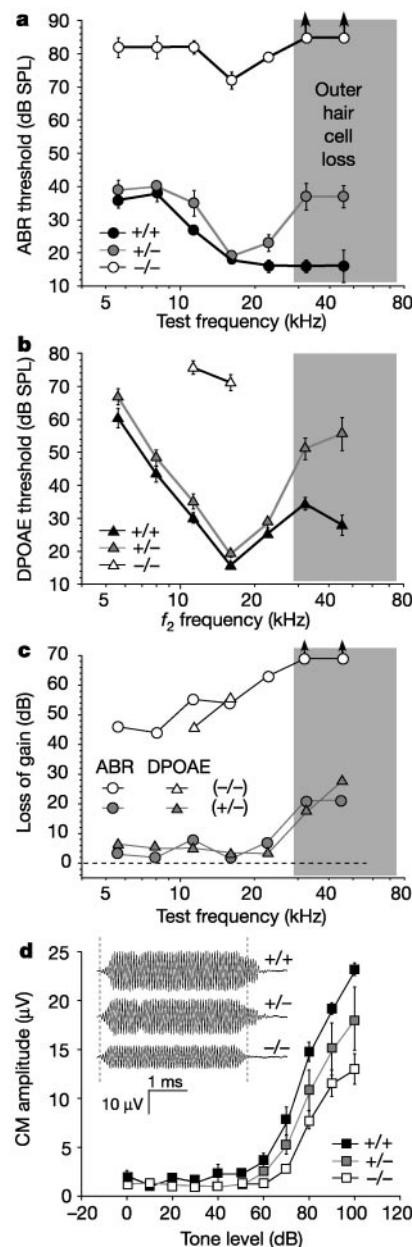


Figure 4 *In vivo* assays of cochlear sensitivity in mutant mice. **a**, Mean ABR thresholds ($\pm$ s.e.m.) for five animals from each group. **b**, Mean DPOAE thresholds ($\pm$ s.e.m.) for ten ears (both ears from five animals) from each group. For f_2 frequencies under 11.3 and over 16 kHz, DPOAE thresholds could not be measured in null mice, because passive system distortion produced a response at levels below the threshold for cochlear distortion. **c**, Contribution of prestin to cochlear amplification, as seen in threshold shifts of null and heterozygous relative to wild-type mice. Data are replotted from panels **a** and **b**. Grey areas in **a–c** correspond to the location where outer hair cell loss was observed in the cochlea of homozygous and heterozygous mice. **d**, Mean CM amplitudes ($\pm$ s.e.m.) compared with sound pressure level for tone bursts at 16 kHz obtained from four animals of each genotype. Inset shows typical CM waveforms for each genotype in response to tones at 80 dB SPL. Dotted lines indicate onset and cessation of the tone burst.

brane voltage is changed under whole-cell voltage clamp (Fig. 3a, top). Outer hair cells isolated from prestin null mice at no point showed measurable motility ($n = 16$). The absolute magnitude of motility in heterozygous cells ($n = 21$) was roughly half that in wild-type cells ($n = 25$) (Fig. 3a–c). Consistent with the histological data (Fig. 2a–c), isolated OHCs also varied in length according to genotype (Fig. 3d, e). To assess whether the prestin deletion affected expression of ion channels, whole-cell currents were measured. No consistent difference was observed between wild-type and prestin null cells, as illustrated by current–voltage curves from OHCs of the apical turn of the cochlea (Fig. 3f).

To test whether prestin, and therefore OHC electromotility, is necessary for normal cochlear mechanical amplification, we obtained three measures of *in vivo* hair cell response from anaesthetized mice at 6–8 weeks of age.

The first measure, auditory brainstem response (ABR), represents synchronous electrical activity generated by neurons in the ascending auditory system and can be recorded from scalp electrodes by averaging responses to short tone bursts. By varying frequency, cochlear function can be assessed throughout the hearing range. In wild-type mice, ABR thresholds vary with frequency (Fig. 4a), from a sound pressure level (SPL) of 35 dB at 5.6 kHz (a relatively low frequency for a mouse) to 15 dB at 45.2 kHz (a relatively high frequency). In prestin null animals, ABR thresholds were 45–65 dB higher than those in wild-type mice, a decrease in sensitivity of two to three orders of magnitude (each 20 dB corresponds to a tenfold change in amplitude). In heterozygous animals, ABR thresholds for responses originating from cochlear regions with all hair cells present (frequencies ≤ 22.6 kHz) were elevated by 1–8 dB.

The second *in vivo* measure, distortion product otoacoustic emissions (DPOAEs), provides a more accurate measure of threshold elevations in heterozygous animals. DPOAEs are mechanical distortions created in the inner ear when two primary tones (f_1 and f_2) are presented. These distortions are amplified by OHCs, and propagated back through the middle ear to the ear canal, where they are measured in the sound pressure waveform. In prestin null mice, system distortion at high sound levels limited measurements to two test frequencies, however the threshold shifts in this region (45–55 dB) were similar to those seen with ABR. In heterozygous mice, DPOAE-based thresholds were 3.1–6.4 dB higher than those in wild-type mice, excluding the highest test frequencies (where OHC loss complicates interpretation). These small differences between the wild-type and heterozygous ears were statistically significant by a two-way analysis of variance (ANOVA; $P = 0.03$).

The final index of *in vivo* measure, the cochlear microphonic (CM), is an electrical response dominated by OHC receptor potentials¹⁸. CM can be recorded through a wire electrode on the surface of the cochlea, and it provides a measure of transducer function in OHCs. The CM amplitudes we observed (Fig. 4d) suggest that mechano-electrical transduction in mutant OHCs is intact. The CM reduction in the ears of heterozygous and prestin null mice is expected, given that CM is dominated by OHC potentials from the basal turn of the cochlea¹⁸, and that as many as half of basal-turn OHCs are missing in the ears of prestin null mice. If OHC transducer currents were absent in the ears of null mice, the CM from remaining IHCs would be less than one-tenth the normal amplitude¹⁸. Note that CM waveforms from the ears of wild-type and heterozygous mice (Fig. 4d) show post-stimulus ‘ringing’, a cochlear ‘echo’ of the type expected in an ‘active’ ear¹⁹; however, in the ear of null mice, CM decays at the cessation of the tone burst, as expected for a passive system.

These results show that prestin is required for both OHC electromotility *in vitro* and normal cochlear amplification *in vivo*. In the absence of prestin, cochlear sensitivity is decreased in a frequency-dependent manner (Fig. 4c) from about 40 dB at 5.6 kHz to more than 60 dB at 22.6 kHz. The magnitudes of these shifts, and their frequency dependence, are consistent with estimates of

cochlear amplifier gain as a function of cochlear frequency. These estimates are made on the basis of the physiologically labile, mechanical nonlinearities in basilar membrane motion^{20–22}, and, more indirectly, on frequency tuning curves of auditory nerve fibres²³. Electromotility and cochlear sensitivity in the prestin heterozygous mouse also fit the simple view that *in vitro* motility and *in vivo* enhancement of cochlear sensitivity are directly coupled. The electromotility of heterozygous cells was roughly half (56%) that of normal cells (Fig. 3b), consistent with the idea that each prestin ‘motor’ makes an equal contribution to OHC length change and the assumption that heterozygous cells express half the number of prestin functional units. If these electrically induced length changes add linearly, *in vivo*, to the mechanical motion of the cochlear partition (at least near threshold values, where cochlear motion increases linearly with sound level²⁴), then a halving of electromotility should be compensated for by a 6-dB increase in sound level. (In the logarithmic decibel scale, 6 dB corresponds to a factor of 2.) Correspondingly, the *in vivo* threshold elevation observed in ears of heterozygous mice was close to 6 dB (Fig. 4c).

The idea that OHC electromotility is the force generator for the cochlear amplifier has long been an attractive hypothesis; however, electromotility is absent in non-mammalian hair cells, which have threshold sensitivities rivaling the mammal²⁵. Furthermore, active movement has been described in hair bundles, which may contribute to mechanical amplification in lower vertebrates^{26,27}. The present data cannot rule out contributions from other cellular motors to the cochlear amplifier in the mouse. Nevertheless, our quantitative analyses of hair cell and cochlear function in prestin mutant mice suggest that no active process other than that provided by prestin-mediated electromotility need be invoked to explain the exquisite mechanical sensitivity of the mammalian cochlea. □

Methods

Generation of prestin mutant mice

Based on gerbil prestin complementary DNA sequence¹⁰, we obtained three overlapping mouse prestin bacterial artificial chromosome (BAC) clones, 399N2, 427N18 and 577A16 (Research Genetics catalogue number 96050). The targeting construct using the NTK Scrambler vector (Stratagene) was transfected into 129/SvEv embryonic stem cells (AB2.2, Stratagene). Four out of 460 clones analysed underwent homologous recombination, identified by Southern blot analysis. The crosses between germline-transmitted heterozygous mice derived from one line yielded offspring with a 1:2:1 (51:105:52) ratio of the wild-type, heterozygous and null genotypes, and all F_2 mice used for further analysis are in a mixed background of 129/SvEv and C57B6/J. PCR primers for genotyping were from the deleted region (5′-GCTTGATGATTGGAGGTGTG-3′ and 5′-CTGAATGATTCTGAAAGTAAGG-3′) and from the targeting vector (5′-CTGTTGTCCAAGTGCTT GCC-3′ and 5′-GATCGCTATCAGGACATAGCG-3′) (Fig. 1d).

Real-time quantitative PCR

We extracted total RNAs from post-natal day 9 (P9) cochleae and designed specific primers and probe for the N-terminal region: 5′-TCGGGCATAAGCACTGGG-3′, 5′-ACGGCTGCCAGCATGG-3′, TaqMan probe, 6FAM-TACTCCAGCTTCCCCAA GGCTTAGCCT-TAMRA. Prestin cDNA was made from a P14 wild-type testis and diluted to generate standard curves for every run of real-time PCR by using the ABI PRISM 7900 Sequence Detection System (Applied Biosystems) in 40 cycles. Cochlear homogenates from each animal were measured in triplicate. The relative quantities of prestin were calculated from standard curves and normalized to 18S (Applied Biosystems catalogue number 4310893E).

Histological analysis and immunocytochemistry

For plastic sections and scanning electron microscopy, mice were perfused with mixed aldehydes. For scanning electron microscopy, whole mounts of cochlear half turns were dried with hexamethyldisilazane, sputter-coated with platinum, and viewed with Philips ESEM XL30. For plastic sections, cochleae were osmicated, decalcified, dehydrated, embedded in plastic, and sectioned at 40 μ m. Each cochlea was reconstructed in three dimensions, cochlear locations were computed and converted to frequency²⁸. For immunohistochemistry, mice were perfused with 10% formalin or 4% paraformaldehyde, and cochleae were extracted, decalcified in EDTA, dehydrated, embedded in paraffin wax, and sectioned at 10 μ m. Sections were incubated in primary antibody overnight followed by either biotinylated secondary antibodies or by fluorescently labelled secondary antibodies.

OHC motility assays

Outer hair cells were isolated from mouse cochleae as described for gerbil OHCs²⁹. Cells were held at -70 mV, and 20-mV steps (-120 – 60 mV) were applied under whole-cell voltage-clamp. Recording pipettes had open-tip resistances of 3–5 M Ω and were filled with

internal solution containing 140 mM KCl, 2 mM MgCl₂, 10 mM EGTA, 10 mM HEPES at pH 7.2. The external solution was Leibovitz's L-15 (Gibco) containing 136 mM NaCl, 5.8 mM NaH₂PO₄, 5.4 mM KCl, 1.3 mM CaCl₂, 0.9 mM MgCl₂ at pH 7.2. Osmolarity was adjusted to 300 mosM l⁻¹. Motility was measured and calibrated using an electro-optical method in which the cell's ciliated pole was imaged through a rectangular slit onto a photodiode²⁹.

In vivo functional assays

For ABR, DPOAE and CM measurements, mice were anaesthetized with xylazine and ketamine. ABRs and DPOAEs were obtained from one set of animals; CMs from a second set. For ABR, needle electrodes were inserted at vertex and pinna. ABR and CM were evoked with 5-ms tone pips (0.5-ms rise-fall, with a cos² envelope, at 35 per s). The response was amplified ($\times 10,000$), filtered (0.1–3 kHz), and averaged with an A/D board in a PC-based data-acquisition system. Sound level was raised in 5-dB steps from 0 to 90 dB SPL. At each level, 1,024 responses were averaged (with stimulus polarity alternated) after 'artefact rejection'. Threshold was determined by visual inspection. For CM a silver-wire electrode was placed on the round window membrane. Responses to alternating pip polarities were subtracted, and the resultant waveform was digitally high-pass filtered to remove residual uncanceled neural potentials. The DPOAE at $2f_1 - f_2$ was recorded in response to two primary tones: f_1 and f_2 , with $f_2/f_1 = 1.2$ and the f_2 level 10 dB lower than the f_1 level. Ear-canal sound pressure was amplified and digitally sampled at 4- μ s intervals. Fast-Fourier transforms were computed from averaged waveforms of ear-canal sound pressure, and the DPOAE amplitude at $2f_1 - f_2$ and surrounding noise floor were extracted. Iso-response contours were interpolated from plots of amplitude versus sound level, performed in 5-dB steps of f_1 level. Threshold is defined as the f_1 level required to produce a DPOAE at 0 dB SPL.

Received 29 May; accepted 12 August 2002; doi:10.1038/nature01059.

Published online 28 August 2002.

- Gold, T. Hearing. II. The physical basis of the action of the cochlea. *Proc. R. Soc. Lond. B* **135**, 492–498 (1948).
- Dallos, P. & Harris, D. Properties of auditory nerve responses in absence of outer hair cells. *J. Neurophysiol.* **41**, 365–383 (1978).
- Brown, M. C., Nuttall, A. L. & Masta, R. I. Intracellular recordings from cochlear inner hair cells: effects of stimulation of the crossed olivocochlear efferents. *Science* **222**, 69–72 (1983).
- Dallos, P. The active cochlea. *J. Neurosci.* **12**, 4575–4585 (1992).
- Brownell, W. E., Bader, C. R., Bertrand, D. & de Ribaupierre, Y. Evoked mechanical responses of isolated cochlear outer hair cells. *Science* **227**, 194–196 (1985).
- Kachar, B., Brownell, W. E., Altschuler, R. & Fex, J. Electrokinetic shape changes of cochlear outer hair cells. *Nature* **322**, 365–368 (1986).
- Ashmore, J. F. A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol.* **388**, 323–347 (1987).
- Ashmore, J. F. *Cochlear Mechanisms* (eds Wilson, J. P. & Kemp, D. T.) 107–116 (Plenum, London, 1989).
- Santos-Sacchi, J. Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J. Neurosci.* **11**, 3096–3110 (1991).
- Zheng, J. et al. Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**, 149–155 (2000).
- Belyantseva, I. A., Adler, H. J., Curi, R., Frolenkov, G. I. & Kachar, B. Expression and localization of prestin and the sugar transporter GLUT-5 during development of electromotility in cochlear outer hair cells. *J. Neurosci.* **20**, RC116 (2000).
- Oliver, D. et al. Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science* **292**, 2340–2343 (2001).
- Forge, A. Structural features of the lateral walls in mammalian cochlear outer hair cells. *Cell Tissue Res.* **265**, 473–483 (1991).
- Dallos, P. & Fakler, B. Prestin, a new type of motor protein. *Nature Rev. Mol. Cell Biol.* **3**, 104–111 (2002).
- Bohne, B. A. & Rabbitt, K. D. Holes in the reticular lamina after noise exposure: implication for continuing damage in the organ of Corti. *Hear. Res.* **11**, 41–53 (1983).
- Holt, J. R. et al. A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells. *Cell* **108**, 371–381 (2002).
- Kros, C. J. et al. Reduced climbing and increased slipping adaptation in cochlear hair cells of mice with Myo7a mutations. *Nature Neurosci.* **5**, 41–47 (2002).
- Dallos, P. & Wang, C. Y. Bioelectric correlates of kanamycin intoxication. *Audiology* **13**, 277–289 (1974).
- Kemp, D. T. Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* **64**, 1386–1391 (1978).
- Johnstone, B. M., Patuzzi, R. & Yates, G. K. Basilar membrane measurements and the travelling wave. *Hear. Res.* **22**, 147–153 (1986).
- Ruggero, M. A. & Rich, N. C. Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration. *Hear. Res.* **51**, 215–230 (1991).
- Sellick, P. M., Patuzzi, R. & Johnstone, B. M. Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique. *J. Acoust. Soc. Am.* **72**, 131–141 (1982).
- Kiang, N. Y. & Moxon, E. C. Tails of tuning curves of auditory-nerve fibers. *J. Acoust. Soc. Am.* **55**, 620–630 (1974).
- Ruggero, M. A. Responses to sound of the basilar membrane of the mammalian cochlea. *Curr. Opin. Neurobiol.* **2**, 449–456 (1992).
- Manley, G. A. Cochlear mechanisms from a phylogenetic viewpoint. *Proc. Natl Acad. Sci. USA* **97**, 11736–11743 (2000).
- Fettiplace, R., Ricci, A. J. & Hackney, C. M. Clues to the cochlear amplifier from the turtle ear. *Trends Neurosci.* **24**, 169–175 (2001).
- Hudspeth, A. J. Mechanical amplification of stimuli by hair cells. *Curr. Opin. Neurobiol.* **7**, 480–486 (1997).
- Ehret, G. *The Auditory Psychobiology of the Mouse* (ed. Willott, J. F.) 169–200 (Charles Thomas, Springfield, Illinois, 1983).
- He, D. Z., Evans, B. N. & Dallos, P. First appearance and development of electromotility in neonatal gerbil outer hair cells. *Hear. Res.* **78**, 77–90 (1994).

Acknowledgements

We thank K. Cullen for technical assistance; T. Curran, B. Fritzsche, C. A. Shera and D. Freeman for comments on the manuscript; and B. Kachar, T. Hasson and P. Gillespie for antibodies. This work is supported in part by NIH grants to M.C.L., Z.Z.H. and J.Z., NIH Cancer Center Support CORE grant, and the American Lebanese Syrian Associated Charities (ALSAC).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.Z.

(e-mail: jian.zuo@stjude.org).

Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning

Avigdor Eldar*†, Ruslan Dorfman*, Daniel Weiss*†, Hilary Ashe‡, Ben-Zion Shilo* & Naama Barkai*†

* Department of Molecular Genetics and † Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel

‡ School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK

Developmental patterning relies on morphogen gradients, which generally involve feedback loops to buffer against perturbations caused by fluctuations in gene dosage and expression¹. Although many gene components involved in such feedback loops have been identified, how they work together to generate a robust pattern remains unclear. Here we study the network of extracellular proteins that patterns the dorsal region of the *Drosophila* embryo by establishing a graded activation of the bone morphogenic protein (BMP) pathway. We find that the BMP activation gradient itself is robust to changes in gene dosage. Computational search for networks that support robustness shows that transport of the BMP class ligands (Scw and Dpp) into the dorsal midline by the BMP inhibitor Sog is the key event in this patterning process. The mechanism underlying robustness relies on the ability to store an excess of signalling molecules in a restricted spatial domain where Sog is largely absent. It requires extensive diffusion of the BMP–Sog complexes, coupled with restricted diffusion of the free ligands. We show experimentally that Dpp is widely diffusible in the presence of Sog but tightly localized in its absence, thus validating a central prediction of our theoretical study.

Graded activation of the BMP pathway subdivides the dorsal region of *Drosophila* embryos into several distinct domains of gene expression. This graded activation is determined by a well-characterized network of extracellular proteins^{2,3}, which may diffuse in the perivitelline fluid⁴ that surrounds the embryo (Fig. 1a). The patterning network is composed of two BMP class ligands (Scw and Dpp), a BMP inhibitor (Sog), a protease that cleaves Sog (Tld) and an accessory protein (Tsg), all of which are highly conserved in evolution and are used also for patterning the dorso-ventral axis of vertebrate embryos⁵. Previous studies have suggested that patterning of the dorsal region is robust to changes in the concentrations of most of the crucial network components. For example, embryos that contain only one functional allele of *scw*, *sog*, *tld* or *tsg* are viable and do not show any apparent phenotype. Misexpression of *scw* or of *tsg* also renders the corresponding null mutants viable^{6–8}.

To check whether robustness is achieved at the initial activation gradient, we monitored signalling directly by using antibodies that recognize specifically an activated, phosphorylated intermediate of

the BMP pathway (pMad)^{9,10}. Prominent graded activation in the dorsal-most eight cell rows was observed for about 1 h, starting roughly 2 h after fertilization at 25 °C (ref. 11 and Fig. 1b). We quantified this activation gradient in heterozygous mutants that were compromised for one of three of the crucial components of the patterning network, Scw, Sog or Tld. Whereas homozygous null mutants that completely lack the normal gene product have a deleterious effect on signalling¹¹, the heterozygotes, which should produce half the amount of the gene product, were indistinguishable from wild type (Fig. 1c). Similarly, overexpression of the Tld protein uniformly in the embryo did not alter the activation profile (Fig. 1c). The activation profile at 18 °C was the same as that at 25 °C (Fig. 1d). This robustness to temperature variations is marked, considering the wide array of temperature dependencies that are observed in this temperature span. By contrast, the profile of pMad was sensitive to the concentration of Dpp¹¹ (Fig. 1d). The dosage sensitivity of Dpp is exceptional among morphogens and is singled out as being haploid-insufficient¹².

No apparent transcriptional feedback, which might account for

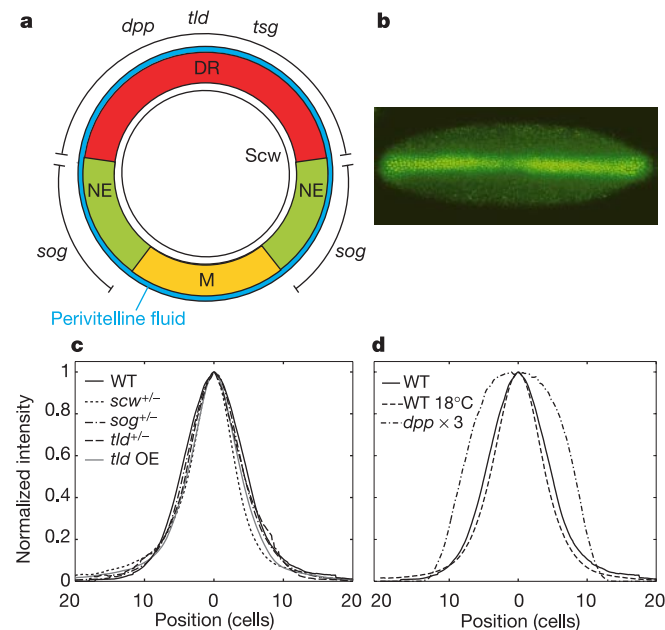


Figure 1 Robustness of the pMad activation profile. **a**, Cross-section of an early *Drosophila* embryo (~2 h after egg lay) showing the three distinct domains of gene expression. Starting at stage 5, the dorsal domain, which comprises about 50 cells, is subdivided to form the amnioserosa and the dorsal ectoderm. Shown are the genes of the patterning network: Scw and Dpp are two activating BMP class ligands; Sog is an inhibitor of both ligands; Tsg is required for Sog inhibition of Dpp; and Tld is a protease that cleaves Sog. Note that *dpp*, *tld* and *tsg* are expressed only in the dorsal region (DR), whereas expression of *sog* is restricted to the neuroectoderm (NE) and *scw* is expressed by all cells. M, mesoderm. **b**, Activation of the BMP pathway, which induces different cell fates in the dorsal region, visualized by antibodies against pMad. The activation profile is shown in a dorsal view of a wild-type embryo at stage 5. Activation is graded and peaks at the dorsal midline. The pattern of pMad widens at the termini of the embryo, possibly owing to edge effects that modify patterning. Our analysis is focused on the centre of the embryo, where the variability between different embryos is limited (~2 cells). **c**, Normalized activation profile of wild-type embryos (averaged over $n = 11$ embryos) compared with that of three sets of heterozygous mutants containing half the amount of Scw ($n = 23$), Sog ($n = 14$) and Tld ($n = 11$). The activation profile of embryos overexpressing the Tld protein uniformly around the embryo, using the Mat α 4-Gal4 driver, is also shown ($n = 33$). All embryos were collected at 25 °C. **d**, Activation profiles at 25 °C and 18 °C ($n = 41$). This temperature variation is biologically significant, as development is about two times slower at 18 °C than at 25 °C. Embryos were collected at different times but at the same developmental stage for the two temperatures. Shown also is the activation profile of an embryo carrying three copies of the *dpp* gene.

the robustness of dorsal patterning, has been identified so far. Robustness should thus be reflected in the design of interactions in the patterning network. To identify the mechanism underlying robustness, we formulated a general mathematical model of the dorsal patterning network. For simplicity, our initial analysis was restricted to a single BMP class ligand (Scw or Dpp), a BMP inhibitor (Sog) and the protease (Tld). The general model accounted for the formation of the BMP–Sog complex, allowed for the diffusion of Sog, BMP and BMP–Sog, and allowed for the cleavage of Sog by Tld, both when Sog is free and when Sog is associated with BMP. Each reaction was characterized by a different rate constant. The three reaction–diffusion equations that define this model are given in the Methods.

We carried out extensive simulations to identify robust networks. At each simulation, a set of parameters (rate constants and protein

Box 1

Mechanism underlying robustness

We consider an idealized patterning network that consists of a single BMP (Scw for simplicity), Sog and Tld. Robustness will be manifested in the steady-state distribution of Scw. We assume that free Scw does not diffuse and that free Sog is not cleaved. The set of reaction–diffusion equations defining this network is obtained from equations 1–3 (Methods), by setting $D_{\text{BMP}} = \alpha_S = k_{-b} = 0$. At steady-state, the system can then be reduced to a single equation:

$$0 = \nabla^2 ([\text{Scw}]^{-1}) - 2l_b^{-2} \quad (1)$$

where $l_b^2 = 2D_S/k_b$, D_S is the diffusion coefficient of Sog and k_b is the binding rate of Sog to Scw. Thus, although free Scw does not diffuse, the system is tuned for providing it with an effective diffusion. Two processes govern this effective diffusion: the shuttling of Scw by Sog from the circumference of the embryo into the dorsal midline, and the degradation of Sog by Tld in the dorsal region. Although both processes depend on the amounts of the respective proteins, those concentrations do not appear in the effective Scw diffusion. The key to this quantitative adjustment is the fact that both processes are mediated by the complex Sog–Scw: only the complex, and not free Scw, can diffuse, and only the complex is subject to degradation by Tld.

The concentrations of the network components, Scw, Sog and Tld could still affect the steady-state activation gradient through the boundary conditions. The solution of equation (1) is given by

$$[\text{Scw}(x)] = \frac{l_b^2}{x^2 + \varepsilon^2} \quad (2)$$

where $x = 0$ at the dorsal midline and ε is an integration coefficient. In general, the value of ε will depend on most parameters of the system, including the concentrations and the production rates of the network components Scw, Sog and Tld, leading to a nonrobust distribution of Scw. Examining equation (2), however, we note that the only way to accommodate high concentrations of Scw is by placing Scw in a small region surrounding the dorsal midline where Sog is absent, in other words, by setting ε to 0. Thus, for large enough concentrations of Scw, ε vanishes and we obtain a robust Scw profile:

$$[\text{Scw}(x)] = \frac{l_b^2}{x^2} \quad (3)$$

for every position of x that is far enough from the dorsal midline, in other words, $x \gg \varepsilon$. Indeed, the concentrations of network components have disappeared from this equation of the activation profile, which reflects the robustness of the system. Finally, we note that the same mechanism generates the gradient of the second BMP (Dpp). We assume, however, that Dpp binds Sog only when the latter is bound to Tsg. Thus, the same formalism is applied but with the molecular entities Dpp, Sog–Tsg (instead of Sog) and the complex Dpp–Sog–Tsg (instead of Scw–Sog). The details and precise conditions necessary for robustness are given in the Supplementary Information.

concentrations) was chosen at random and the steady-state activation profile was calculated by solving equations (1) to (3) numerically. A set of three perturbed networks representing heterozygous situations was then generated by reducing the gene dosages of *sog*, *tld* or the BMP class ligand by a factor of two. The steady-state activation profiles defined by those networks were solved numerically and compared with the initial, nonperturbed network. A threshold was defined as a given BMP value (corresponding to the value at a third of the dorsal ectoderm in the nonperturbed network). The extent of network robustness was quantified by measuring the shift in the threshold for all three perturbed networks. Over 66,000 simulations were carried out, with each of the nine parameters allowed to vary over four orders of magnitude.

As expected, in most cases (97.5%) the threshold position in the perturbed networks was shifted by a large extent (>50%; see Fig. 2a). In most of those nonrobust cases, the BMP concentration was roughly uniform throughout the dorsal region (Fig. 2c). By contrast, Sog was distributed in a concentration gradient with its

minimum in the dorsal midline, defining a reciprocal gradient of BMP activation. Thus, the key event in this nonrobust patterning mechanism was the establishment of a concentration gradient of Sog, which was governed by diffusion of Sog from its domain of expression outside the dorsal region, coupled with its cleavage by Tld inside the dorsal region. Although such a gradient has been observed¹³, it is also compatible with other models (see below).

We identified a small class of networks (198 networks, 0.3%) in which a twofold reduction in the amounts of all three genes resulted in a change of less than 10% in the threshold position (see Fig. 2b). Notably, in all of these robust cases, BMP was redistributed in a sharp concentration gradient that peaked in the dorsal midline (Fig. 2c). In addition, this concentration gradient decreased as a power-law distribution with an exponent $n = 2$, which indicated the uniqueness of the robust solution (Fig. 2d). In these cases, Sog was also distributed in a graded manner in the dorsal region (data not shown). Analysis of the reaction rate constants of the robust networks showed a wide range of possibilities for most parameters. But two restrictions were apparent and defined the robust network design. First, in the robust networks the cleavage of Sog by Tld was facilitated by the formation of the complex Sog-BMP (Fig. 2f). Second, the complex BMP-Sog was broadly diffusible, whereas free BMP was restricted (Fig. 2e).

To identify how robustness is achieved, we considered an idealized network by assuming that free Sog is not cleaved and that free BMP does not diffuse. The steady-state activation profile defined by this network can be solved analytically (Box 1), which reveals the two aspects that are crucial for ensuring robustness. First, the BMP-Sog complex has a central role, by coupling the two processes that establish the activation gradient: BMP diffusion and Sog degradation. This coupling leads to a quantitative buffering of perturbations in gene dosage. Second, restricted diffusion of free BMP enables the system to store excess BMP in a confined spatial domain where Sog is largely absent. Changes in the concentration of BMP

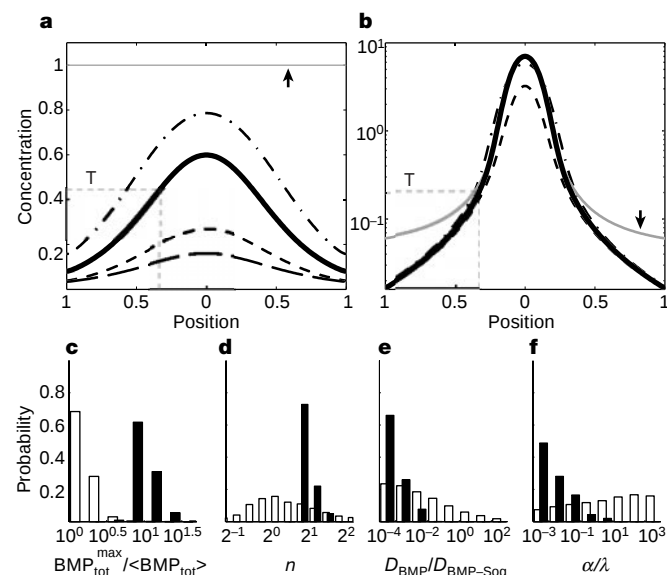


Figure 2 Patterning mechanism emerges from the features of the robust networks. The general model of the dorsal patterning network was solved numerically for 66,000 different choices of parameters, with each parameter ranging over four orders of magnitude. **a**, A typical, nonrobust network. The profile of free BMP (unbroken curves) is shown for the nonperturbed network and for three perturbed networks representing heterozygotes for *sog*, *tld* and BMP (see Fig. 1c for a key to the lines). The total concentration of BMP (free plus Sog-associated) is indicated by the grey line (arrow). The broken grey line (T) indicates the threshold where robustness was measured. **b**, A typical robust system. **c–f**, Statistical distribution of various features in the robust (black) and nonrobust (white) networks. The analysis was restricted to the 22,000 networks that showed at least twofold spatial variation of free BMP concentration. Each feature was calculated for each of the networks, and the histograms were normalized to account for the different numbers of robust (198) and nonrobust (22,000) networks. **c**, The extent of BMP confinement to the dorsal midline was quantified by measuring the ratio between total BMP concentration (free plus Sog-associated) at the centre and its average concentration. In all of the robust cases, a high ratio (>10) was observed. By contrast, the low ratios observed in the nonrobust cases indicated that BMP was distributed approximately uniformly. **d**, The steady-state profile of free BMP was fit to a power-law distribution, x^{-n} . Nearly all robust profiles corresponded to $n = 2$, indicating the uniqueness of the robust solution. By contrast, a wide range of exponents were found for the nonrobust solution. The fitting error vanished in the robust cases, but was high in the nonrobust cases (not shown). **e**, Ratio between the diffusion coefficient of free BMP (D_{BMP}) and the complex BMP-Sog ($D_{\text{BMP-Sog}}$). Note that complex formation significantly enhances BMP diffusion in all robust cases. **f**, Ratio of the degradation rate of free Sog (α) to that of BMP-associated Sog (λ). Note that complex formation greatly enhances Sog degradation in all robust cases.

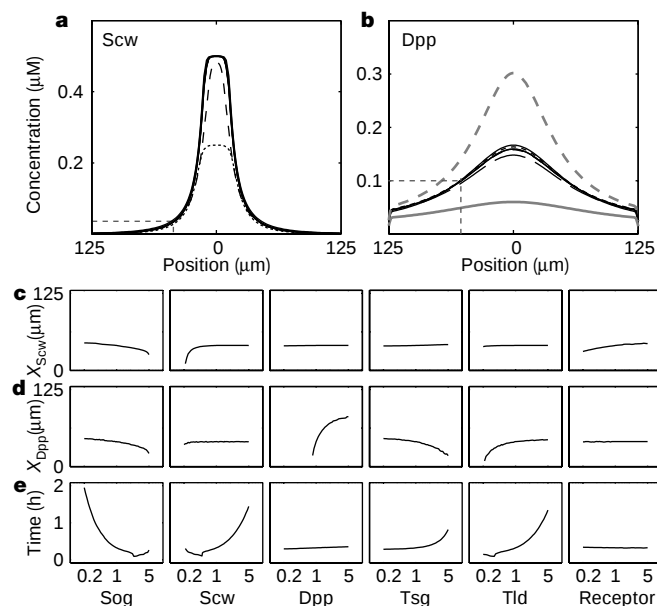


Figure 3 Properties of the robust model. **a, b**, The steady-state concentrations of receptor-bound Scw and Dpp. Six altered systems were obtained from a reference system by reducing to half the amounts of *scw*, *sog*, *tld* (see Fig. 1c for a key to the lines), *dpp* (unbroken grey line), Scw receptor (dotted line), *tsg* (superimposed on the solid line) and by increasing by 50% the amount of *dpp* (grey dashed line). The level and position of the threshold used in **c** and **d** are indicated. **c, d**, The positions of the activation thresholds of Scw (**c**) and Dpp (**d**) for a series of altered systems, obtained by changing a single parameter by the indicated fold amount. **e**, Time to reach steady state as a function of the fold change in parameter values. Time is initiated with the onset of ligand production.

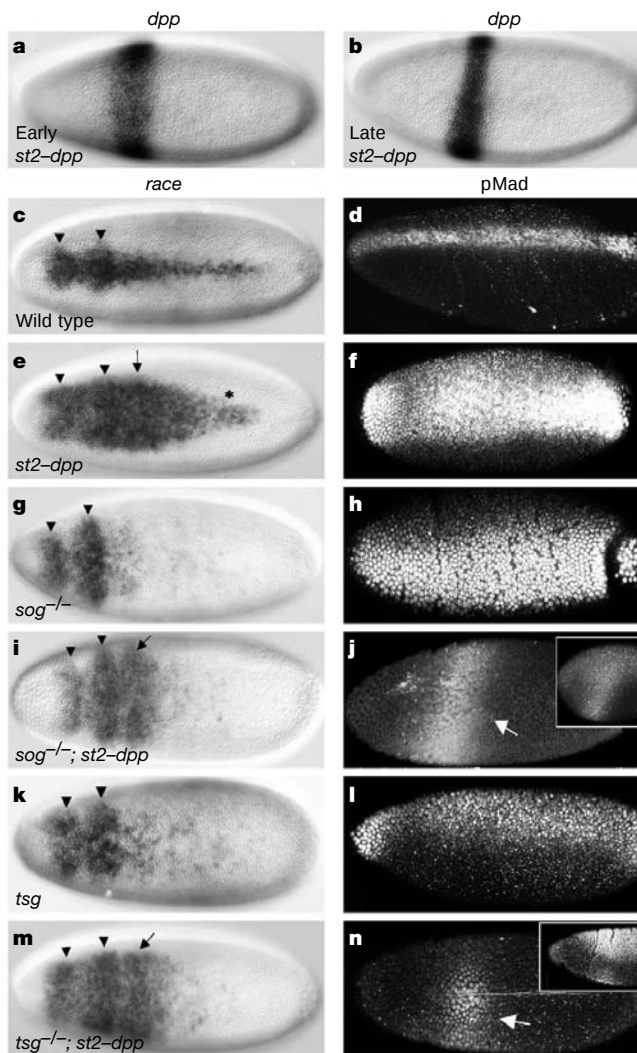


Figure 4 Dpp diffusion requires Sog and Tsg. Dpp was expressed orthogonal to the dorso-ventral axis by the *eve st2* enhancer in a stripe of about 12 cells at early cleavage cycle 14 (**a**), which refines to about 6 cells by late cycle 14 (lateral views, **b**). Note that ectopic amounts of *dpp* were greater than the endogenous transcripts in the dorsal region, which under these exposures were not visible. **c**, Normal pattern of *race* in wild-type embryos. Note the AP bias in the anterior positions (arrowheads). **d**, Normal pattern of pMad in wild-type embryos. **e**, Expression of *st2-dpp* in wild-type embryos leads to an expanded expression of *race*, which forms a wedge shape with its broadest region at the position of *st2-dpp* (arrow) and ranges about 25 cells posteriorly (asterisk). The normal pattern of *race* shows an AP bias at the anterior positions (arrowheads). **f**, *st2-dpp* in wild-type embryos also leads to a broader dorsal distribution of pMad, which extends throughout the AP axis. **g**, In *sog*^{-/-} embryos, the anterior *race* domain expands (arrowheads), whereas posterior expression is diminished and detected only sporadically as a punctate staining. Some expansion of *race* was observed in *sog*^{+/-} embryos, probably due to the higher threshold of *race* induction compared with pMad. **h**, In *sog*^{-/-} embryos, pMad expands in the dorsal domain, but does not reach the ventral domain. **i**, Expression of *st2-dpp* in *sog*^{-/-} embryos leads to a corresponding stripe of *race* of about 10–12 cells (arrow). **j**, *st2-dpp* in *sog*^{-/-} embryos generates a corresponding stripe of pMad (arrow), which extends to the ventral region. The ventral expansion of pMad versus the restricted dorsal expansion of *race* in embryos of the same genotype indicates that a lower threshold of activation can induce pMad even in the ventral domain, which is devoid of endogenous Dpp. Main view is ventral, inset is lateral. **k**, **l**, The patterns of *race* and pMad in *tsg*^{-/-} embryos are similar to *sog*^{-/-} embryos. **m**, **n**, *st2-dpp* in *tsg*^{-/-} generates a corresponding stripe of both *race* and pMad, similar to that observed on expression of *st2-dpp* in *sog*^{-/-} embryos. Main view in **n** is ventral, inset is lateral. Except where stated otherwise, a dorsal view is shown with anterior to the left.

alter the BMP profile close to the dorsal midline but do not change its distribution in most of the dorsal region (Box 1).

We next examined the complete system, comprising Sog, Tld, Tsg, both Scw and Dpp, and their associated receptors (Supplementary Information). Two additional molecular assumptions were required to ensure the robustness of patterning. First, Sog can bind and capture the BMP class ligands even when the latter are associated with their receptors. Second, Dpp can bind Sog only when the latter is bound to Tsg. Indeed, it has been shown that, whereas Sog is sufficient for inhibiting Scw, both Tsg and Sog are required for inhibiting Dpp^{6,14,15}. This last assumption implies that Tsg functions to decouple the formation of the Scw gradient from the parallel generation of the Dpp gradient, ensuring that Scw and Dpp are transported to the dorsal midline independently by two distinct molecular entities (Supplementary Information).

The complete model was solved numerically for different choices of rate constants. In particular, we assessed the effect of twofold changes in gene dosage. The steady-state activation profiles can be superimposed, indicating the robustness of the system (Fig. 3a, b). In addition, with the exception of Dpp, the expression of all other crucial network components can be altered by at least an order of magnitude before an effect on the position of a given threshold is observed (Fig. 3c, d). In the model, the lack of robustness to Dpp stems from its insufficient dosage. Note that the time taken to reach steady state is sensitive to these concentrations of protein (Fig. 3e). For the wide range of parameters that we have used, however, the adjustment time does not exceed the patterning time. Flexible adjustment time thus facilitates the buffering of quantitative perturbations.

As discussed above, our analysis identified two principle molecular features that are essential for robust network design: first, free Sog is not cleaved efficiently—an assumption that is supported by the *in vitro* finding that Sog cleavage by Tld requires BMP^{6,16}; second, the diffusion of free BMP is restricted. This is the central prediction of our theoretical study, namely, that Scw diffusion requires Sog, whereas Dpp diffusion requires both Sog and Tsg. Although several reports suggest that in wild-type embryos both Dpp and Scw are widely diffusible^{6,17}, their ability to diffuse in a *sog* or *tsg* mutant background has not been examined as yet.

To monitor the diffusion of Scw or Dpp, we used the *even-skipped* (*eve*) stripe-2 enhancer (*st2*) to misexpress Dpp or Scw in a narrow stripe perpendicular to the normal BMP gradient. In transgenic embryos, *dpp* or *scw* RNA was detected in a stripe just posterior to the cephalic furrow. Initially the stripe was about 12 cells wide at early cleavage cycle 14, but refined rapidly to about 6 cells by late cycle 14 (Fig. 4a, b). The *st2-dpp* and *st2-scw* embryos were viable, despite the high expression of these proteins as compared with their endogenous counterparts.

The activation of the BMP pathway was monitored either by staining for pMad or by following dorsal expression of the target gene *race*, which requires high activation. Scw is a less potent ligand than is Dpp. This experimental setup could not be used to study Scw diffusion properties because expressing *st2-scw* did not alter the pattern of pMad or *race* expression in wild-type or *sog*^{-/-} embryos (data not shown). By contrast, expression of *st2-dpp* led to an expansion of both markers in a region that extends far from the *st2* expression domain, indicating a wide diffusion of Dpp in a wild-type background (compare Fig. 4c, d with 4e, f). Conversely, on expression of *st2-dpp* in *sog*^{-/-} or in *tsg*^{-/-} embryos, both markers were confined to a narrow stripe in the *st2* domain (compare Fig. 4g, h and k, l with 4i, j and m, n, respectively). The width of this stripe was comparable to that of *st2-dpp* expression, ranging from 6 to 12 cells, indicating that Dpp does not diffuse from its domain of expression in the absence of Sog or Tsg. Taken together, these results show that both Sog and Tsg are required for Dpp diffusion, as predicted by the theoretical analysis.

The computation ability of biochemical networks is striking when one considers that they function in a biological environment

where the amounts of the network components fluctuate, the kinetics is stochastic, and sensitive interactions between different computation modules are required. Studies have examined the effect of these properties on cellular computation mechanisms^{18–20}, and robustness has been proposed to be a ‘design principle’ of biochemical networks^{18,21}. We have shown the applicability of this principle to morphogen gradient patterning during early development. Quantitative analysis can be used to assess rigorously the robustness of different patterning models and to exclude incompatible ones. The remaining, most plausible model points to crucial biological assumptions and serves to postulate the central feedback mechanisms. Applying the same modelling principles to other systems might identify additional ‘design principles’ that underlie robust patterning by morphogen gradients in development. □

Methods

Fly strains

We used the following strains: *screw*¹², *sog*⁶, *tolloid*², *tolloid*², UAS-*tld* (provided by M. O’Connor), *Matα4-Gal4 VP16* (provided by D. St. Johnston) and *tsg*^{XB56} (provided by L. Marsh). For altering the number of copies of *dpp*, we used the strain *dpp*¹⁸⁴⁶ *sp cn bw/ CyO 23P[dpp⁺]* (provided by S. Roth), which was crossed to wild-type flies. The mutant chromosomes were maintained over a balancer chromosome. When each strain is crossed to itself, two-thirds of the embryos not showing the null phenotype should be heterozygotes for *scw* and *tld*, and a third for *sog*. The *st2-dpp* strain has been described¹⁷. We constructed *st2-scw* by inserting a *scw* cDNA fragment into plasmid 22FPE (ref. 22).

Antibodies and staining

Rabbit antibodies against pMad were kindly provided by P. ten Dijke. Freshly collected embryos were fixed in 7% formaldehyde. The remaining steps of staining were done according to standard procedures. We used Cy2-conjugated secondary antibodies against rabbit IgG (Jackson ImmunoResearch).

Data analysis

The dorso-ventral activation profile was quantified by using image processing tools written in Matlab. The mean intensity was measured at the middle of the anterior-posterior (AP) axis in a swath that was 20 cells wide in the direction of the AP axis. We then averaged the results over the indicated number of embryos.

Numerical simulations

For the general patterning model we considered a single BMP, which is denoted here as Scw. The model was defined by the following set of reaction–diffusion equations:

$$\frac{\partial[\text{Sog}]}{\partial t} = D_S \nabla^2 [\text{Sog}] - k_b [\text{Sog}][\text{Scw}] + k_{-b} [\text{Sog-Scw}] - \alpha [\text{Tld}][\text{Sog}] \quad (1)$$

$$\frac{\partial[\text{Scw}]}{\partial t} = D_{\text{BMP}} \nabla^2 [\text{Scw}] - k_b [\text{Sog}][\text{Scw}] + \lambda [\text{Tld}][\text{Sog-Scw}] + k_{-b} [\text{Sog-Scw}] \quad (2)$$

$$\frac{\partial[\text{Sog-Scw}]}{\partial t} = D_C \nabla^2 [\text{Sog-Scw}] + k_b [\text{Sog}][\text{Scw}] - k_{-b} [\text{Sog-Scw}] - \lambda [\text{Tld}] \times [\text{Sog-Scw}] \quad (3)$$

The equations were solved in the region $-1 < x < 1$. The parameters in this model include the diffusion coefficients of Sog, Scw and the complex Scw–Sog, (D_S , D_{BMP} and D_C), binding and unbinding of the Scw–Sog complex (k_b and k_{-b}), cleavage of Sog by Tld (α when Sog is free, λ when Sog is associated with Scw), a constant flux of Sog on the boundaries (η_s) and the total Scw concentration ($[\text{Scw}]_{\text{av}}$). D_C is $D_{\text{BMP-Sog}}$ in Fig. 2e. We solved the equations for 66,000 different sets of random parameters. Each parameter was allowed to vary over four orders of magnitude. The parameters defining the centre of this distributions are (in arbitrary units): $D_S = I_1 = [\text{Scw}]_{\text{av}} = 1$, $D_{\text{BMP}} = 0.1$, $D_C = 1$, $k_b = 10$, $k_{-b} = 1$, $\lambda [\text{Tld}] = 10$, $\alpha [\text{Tld}] = 10$, $\eta_s = 10$. Equations were solved with Matlab. Each run took less than 1 min. The parameters of the systems shown in Fig. 2 are specified in the Supplementary Information.

For the parameters of the full model (Fig. 3), we chose diffusion rates that reflected the rapid *in vivo* patterning time and corresponded to the measured diffusion time in the perivitelline fluid⁴. This measured diffusion coefficient is similar to that of green fluorescent protein (GFP) in water²³. It is possible that mixing processes in the perivitelline fluid contribute to the equilibration process. For simplicity, we approximate such processes as an effective diffusion. This approximation does not affect our conclusions. No biochemical data restricting the values of the other parameters are available. The parameters of the reference system are within the realistic biochemical range and obey the robustness conditions. The parameter choice is specified and rationalized in detail in the Supplementary Information.

Received 11 April; accepted 27 July 2002; doi:10.1038/nature01061.

- Freeman, M. Feedback control of intercellular signalling in development. *Nature* **408**, 313–319 (2000).
- Podos, S. D. & Ferguson, E. L. Morphogen gradients: new insights from DPP. *Trends Genet.* **15**, 396–402 (1999).
- Rafferty, L. A. & Sutherland, D. J. TGF- β family signal transduction in *Drosophila* development: from Mad to Smads. *Dev. Biol.* **210**, 251–268 (1999).

- Stein, D., Roth, S., Vogelsang, E. & Nusslein-Volhard, C. The polarity of the dorsoventral axis in the *Drosophila* embryo is defined by an extracellular signal. *Cell* **65**, 725–735 (1991).
- DeRobertis, E. M. & Sasai, Y. A common plan for dorsoventral patterning in Bilateria. *Nature* **380**, 37–40 (1996).
- Nguyen, M., Park, S., Marques, G. & Arora, K. Interpretation of a BMP activity gradient in *Drosophila* embryos depends on synergistic signaling by two type I receptors, SAX and TKV. *Cell* **95**, 495–506 (1998).
- Arora, K., Levine, M. S. & O’Connor, M. B. The screw gene encodes a ubiquitously expressed member of the TGF- β family required for specification of dorsal cell fates in the *Drosophila* embryo. *Genes Dev.* **8**, 2588–2601 (1994).
- Mason, E. D., Williams, S., Grotendorst, G. R. & Marsh, J. L. Combinatorial signaling by Twisted Gastrulation and Decapentaplegic. *Mech. Dev.* **64**, 61–75 (1997).
- Persson, U. *et al.* The L45 loop in type I receptors for TGF- β family members is a critical determinant in specifying Smad isoform activation. *FEBS Lett.* **434**, 83–87 (1998).
- Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. Hedgehog creates a gradient of DPP activity in *Drosophila* wing imaginal discs. *Mol. Cell* **5**, 59–71 (2000).
- Dorfman, R. & Shilo, B.-Z. Biphasic activation of the BMP pathway in the *Drosophila* embryonic dorsal region. *Development* **128**, 965–972 (2001).
- Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**, 807–822 (1993).
- Srinivasan, S., Rashka, K. E. & Bier, E. Creation of a Sog morphogen gradient in the *Drosophila* embryo. *Dev. Cell* **2**, 91–101 (2002).
- Neul, J. L. & Ferguson, E. L. Spatially restricted activation of the SAX receptor by SCW modulates DPP/TKV signaling in *Drosophila* dorsal–ventral patterning. *Cell* **95**, 483–494 (1998).
- Ross, J. J. *et al.* Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature* **410**, 479–483 (2001).
- Marques, G. *et al.* Production of a DPP activity gradient in the early *Drosophila* embryo through the opposing actions of the SOG and TLD proteins. *Cell* **91**, 417–426 (1997).
- Ashe, H. L., Mannervik, M. & Levine, M. Dpp signaling thresholds in the dorsal ectoderm of the *Drosophila* embryo. *Development* **127**, 3305–3312 (2000).
- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Barkai, N. & Leibler, S. Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
- von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. The segment polarity network is a robust developmental module. *Nature* **406**, 188–192 (2000).
- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular biology. *Nature* **402**, C47–C52 (1999).
- Kosman, D. & Small, S. Concentration-dependent patterning by an ectopic expression domain of the *Drosophila* gap gene *knirps*. *Development* **124**, 1343–1354 (1997).
- Swaminathan, R., Hoang, C. P. & Verkman, A. S. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP–S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys. J.* **72**, 1900–1907 (1997).

Supplementary Information accompanies the paper on Nature’s website (<http://www.nature.com/nature>).

Acknowledgements

We thank P. ten Dijke for the pMad antibody; L. Marsh, M. O’Connor, S. Roth, D. St. Johnston and the Umea and Bloomington Fly Centers for strains; and S. Leibler and S. Roth for comments and criticism. This work was funded by the Israel Science Foundation (B.-Z.S.) and the Israel Science Foundation and the Minerva Foundation (N.B.). H.A. is a Lister Institute Research Fellow. B.-Z.S. is the incumbent of the Hilda and Cecil Lewis professorial chair in Molecular Genetics. N.B. is the incumbent of the Sorella and Henry Shapiro career development chair.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.B. (e-mail: naama.barkai@weizmann.ac.il).

Molecular basis of seasonal time measurement in *Arabidopsis*

Marcelo J. Yanovsky & Steve A. Kay

The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA

Several organisms have evolved the ability to measure daylength, or photoperiod, allowing them to adjust their development in anticipation of annual seasonal changes. Daylength measurement requires the integration of temporal information, provided by the circadian system, with light/dark discrimination, initiated by specific photoreceptors. Here we demonstrate that in *Arabidopsis* this integration takes place at the level of CONSTANS

(CO)¹ function. CO is a transcriptional activator that accelerates flowering time in long days, at least in part by inducing the expression of *FLOWERING LOCUS T* (*FT*)^{2–5}. First, we show that precise clock control of the timing of CO expression, such that it is high during daytime only in long days, is critical for daylength discrimination. We then provide evidence that CO activation of *FT* expression requires the presence of light perceived through cryptochrome 2 (*cry2*) or phytochrome A (*phyA*). We conclude that an external coincidence mechanism, based on the endogenous circadian control of CO messenger RNA levels, and the modulation of CO function by light, constitutes the molecular basis for the regulation of flowering time by daylength in *Arabidopsis*.

The molecular genetic dissection of flowering time in *Arabidopsis* has identified several photoreceptors and light signalling proteins, as well as some putative clock components, that are essential for an appropriate photoperiodic response⁶. However, we still lack an adequate understanding of how temporal and light environment information are integrated at the molecular level to allow daylength regulation of flowering time. It has recently been proposed that this integration may take place at the level of CO⁷. The expression of CO is regulated by the clock and photoperiod, with high levels of CO mRNA occurring during daytime in long days but not in short days. Levels of *FT* mRNA, a direct target of CO, peak in long days at dusk when there is a coincidence between high levels of CO mRNA and the illuminated part of the day⁷, so CO activity may be regulated by light. In support of the idea that clock control of CO expression may be relevant for daylength discrimination, it has been shown that changes in the levels of CO mRNA are responsible for the alterations in flowering time observed in the mutants *gigantea* (*gi*), *late elongated hypocotyl* (*lhy*) and probably also in *early flowering 3* (*elf3*)⁷, all of which have circadian defects^{8–10}. However, these mutants also have light-dependent phenotypes^{8,11–13} making it impossible to rule out that their effects on CO mRNA levels are independent of the circadian clock. Furthermore, proving that clock control of CO expression is important for photoperiodic time measurement requires evidence that precise timing (rather than the overall level) of CO expression is needed for correct daylength interpretation.

The clock mutant *toc1-1* constitutes an ideal system to analyse the relevance of the circadian control of CO expression for the photoperiodic regulation of flowering time. TOC1 is a pseudo-response regulator closely associated with the central oscillator that, when mutated, causes early flowering under short days and reduced

Table 1 Flowering time of wild-type plants and mutants

Genotype	Light/dark cycle (hours)	Leaf number
Wild-type <i>Ler</i>	8 L: 16 D	22.1 ± 0.9
<i>co-2</i>	8 L: 16 D	23.3 ± 0.9
<i>ft-2</i>	8 L: 16 D	24.5 ± 0.8
<i>toc1-1</i>	8 L: 16 D	8.7 ± 0.5
<i>toc1-1 co-2</i>	8 L: 16 D	18.0 ± 0.6
<i>toc1-1 ft-2</i>	8 L: 16 D	21.5 ± 0.8
Wild-type <i>Ler</i>	10 L: 20 D	7.4 ± 0.2

Flowering time was measured as number of rosette leaves at bolting. Data are mean ± s.e.m. of 10–16 plants, grown in short days of 24 (8 h light/16 h dark) or 30 h (10 h light/16 h dark) of total duration.

daylength sensitivity (particularly in the Landsberg *erecta* ecotype), in addition to shortening the circadian period of all the clock outputs analysed^{14–17}. The *toc1-1* mutant lacks any obvious light-dependent morphological phenotype and its circadian defect persists under a wide range of light conditions, even in complete darkness^{14,15}. More importantly, the early flowering phenotype of the short period mutant *toc1-1* is observed when the plants are grown under short days of 24 h (that is, 8 h light/16 h dark (8 L: 16 D)), but not under short days of 21 h (7 L: 14 D), an environmental cycle that matches the endogenous period of this mutant¹⁴. This indicates that the photoperiodic defect of *toc1-1* is due entirely to its circadian defect.

Here we have analysed the expression of CO in the *toc1-1* mutant and in wild-type plants grown under short and long days of 24 and 21 h of total duration, to test whether the photoperiodic defect of *toc1-1* is the result of incorrect modulation of phase and/or waveform of CO expression. In wild-type plants grown under short days CO mRNA levels accumulated mostly during the dark period (Fig. 1a, b), whereas under long days CO expression was moderate to high also at dawn and dusk (Fig. 1c, d). In *toc1-1* mutants grown under short days of 24 h, the phase of CO expression was significantly advanced, leading to a coincidence between relatively high levels of CO mRNA and the illuminated part of the day at dusk (Fig. 1a). No shift in CO expression was observed in *toc1-1* seedlings grown under short days of 21 h (Fig. 1b), a condition in which the flowering-time phenotype of the mutant is similar to that of wild-type seedlings¹⁴. We note that the waveform and levels of expression of *LHY* and *GI* were not affected in the *toc1-1* mutant under short days of 24 h (Supplementary Information Fig. 1), indicating that the flowering-time defect of this mutant was not due to a generalized alteration in the expression of clock-regulated flowering-time genes.

The *co-2* mutation by itself does not affect flowering time under

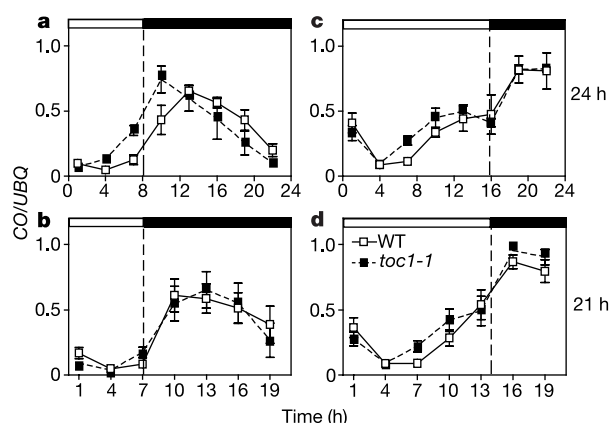


Figure 1 CO expression in the *toc1-1* mutant and in wild-type plants. **a, b**, Plants grown in short days of 24 h (8 L: 16 D) or 21 h (7 L: 14 D), respectively. **c, d**, Plants grown in long days of 24 h (16 L: 8 D) and 21 h (14 L: 7 D), respectively. Data are mean ± s.e.m. of 3–4 samples from completely independent experiments. White and dark bars above traces represent light and dark periods respectively. WT, wild type.

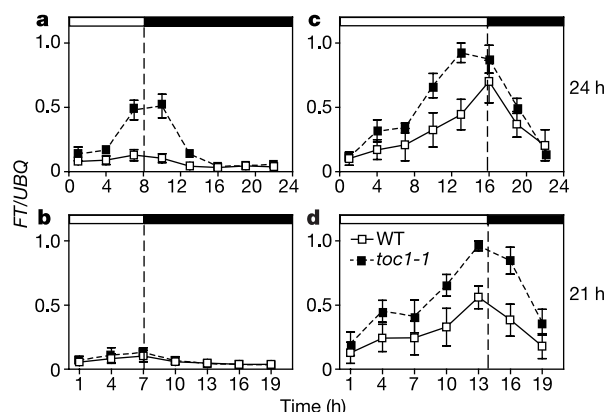


Figure 2 FT expression in the *toc1-1* mutant and in wild-type plants. **a, b**, Plants grown in short days of 24 h (8 L: 16 D) or 21 h (7 L: 14 D), respectively. **c, d**, Plants grown in long days of 24 h (16 L: 8 D) and 21 h (14 L: 7 D), respectively. Data are mean ± s.e.m. of 3–4 samples from completely independent experiments. White and dark bars above traces represent light and dark periods respectively.

short days of 24 h (ref. 1), but it delayed flowering time when present in a *toc1-1* background (Table 1). This effect was not due to a reversion of the *toc1-1* circadian phenotype (Supplementary Information Fig. 2). Because the *toc1-1* mutation does not affect the overall levels of *CO* mRNA, but only the phase angle of its expression, these results indicate that the change in the waveform of *CO* expression was responsible for the flowering-time defect of the *toc1-1* mutant and that the appropriate circadian control of the timing of *CO* expression is critical for daylength discrimination. The fact that the *toc1-1 co-2* double mutant still flowered slightly earlier than wild-type seedlings (Table 1) suggests the involvement of other genes with *CO*-like function.

Molecular and genetic evidence indicates that *CO* accelerates flowering time in long days through a direct activation of *FT* expression²⁻⁵. We therefore analysed whether the change in *CO* phase-angle of expression decreased flowering time in the *toc1-1* mutant in short days by a similar mechanism. In wild-type plants *FT* expression was barely detectable under short days (Fig. 2a, b), but it accumulated during the illuminated part of the day under long days (Fig. 2c, d). In the *toc1-1* mutant, *FT* mRNA levels accumulated not only in long days (Fig. 2c, d) but also during the illuminated part of the day in short days of 24 h (Fig. 2a). On the other hand, very low levels of *FT* mRNA were observed in *toc1-1* seedlings grown under short days of 21 h (Fig. 2b), a condition in which the waveform of *CO* expression in the mutant was quite similar to that seen in the wild type (Fig. 1b). The *ft-2* mutation alone only slightly delays flowering time in short days⁴, but caused a dramatic delay in a *toc1-1* background (Table 1), indicating that the increased levels of *FT* mRNA observed in *toc1-1* are indeed the molecular basis for its early flowering phenotype.

If the effects of *toc1-1* on *CO* and *FT* expression were only due to the period length defect of the mutant, similar changes should be observed in wild-type plants grown under short days of 30 h (10 L: 20 D). In support of this, *CO* expression was shifted towards daytime in wild-type plants grown under cycles of 10 L: 20 D (Fig. 3a). This shift in *CO* expression correlated not only with an upregulation of *FT* mRNA levels (Fig. 3b) but also with a reduction in flowering time (Table 1). The higher levels of *FT* mRNA observed in plants grown under cycles of 10 L: 20 D compared to those of 8 L: 16 D were due to the difference in the total duration of the cycle and not to the longer light interval, because *FT* mRNA levels in plants grown under 10 L: 14 D were low and similar to those present in 8 L: 16 D (data not shown).

FT mRNA levels in plants expressing *CO* from a strong and constitutive promoter are higher than those of wild-type seedlings, but are still rhythmic (peaking close to dusk) when the plants are grown in long days⁷. This could be due to either light regulation of *CO* function and/or *CO*-independent circadian regulation of *FT* expression. To specifically assess the effect of light we analysed *FT* expression in *CO*-overexpressing plants grown in short days and then kept for 24 h under constant light or dark conditions. *FT* mRNA levels increased and remained high during all the exper-

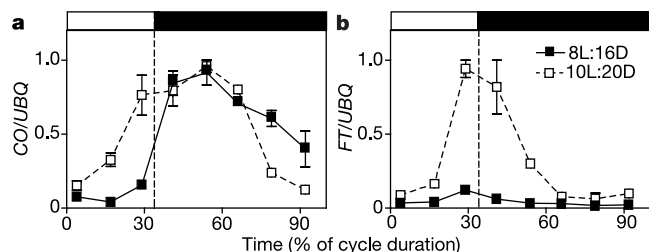


Figure 3 *CO* and *FT* expression in wild-type plants. Plants were grown under short days of 24 h (8 L: 16 D) of total duration or under short days of 30 h (10 L: 20 D) of total duration. Data are mean $\pm$ s.e.m. of two independent samples. Time is shown as percentage of cycle duration: hours from lights on divided by cycle duration (in hours) multiplied by 100.

imental period in plants transferred to constant light, but slowly decreased in those transferred to darkness (Fig. 4a), despite the fact that *CO* expression was constant and similar in both conditions (Fig. 4b). The conditional nature of *FT* mRNA accumulation in plants overexpressing *CO* provides strong evidence that *CO* regulation of *FT* mRNA levels is light dependent. The analysis was done for only 24 h, so we cannot exclude the existence of an additional *CO*-independent circadian regulation of *FT* expression.

The blue-light photoreceptor *cry2* is a likely candidate for the light-regulated promotion of *FT* expression by *CO*, because *cry2* loss-of-function mutants flower late in long days but not in short days¹⁸. In agreement with this hypothesis we found that, compared to wild-type seedlings grown in long days, *FT* mRNA levels are greatly reduced in *cry2* mutants (Fig. 4c), whereas, as recently reported⁷, no difference is observed in the levels or phase-angle of *CO* expression (Fig. 4d). Another photoreceptor mutant impaired in daylength perception is *phyA*¹⁹. This mutant flowers late when short days of fluorescent light are extended with several hours of incandescent light, which is rich in far-red light (that is, the part of the spectrum that most efficiently activates *phyA*). The *phyA* mutant also displayed a dramatic reduction in *FT* mRNA levels when grown under short days extended with incandescent light (Fig. 4e). However, this mutant also showed an abnormal waveform of expression and slightly reduced levels of *CO* mRNA (Fig. 4f), suggesting that a minor part of its effect on *FT* expression might be due to changes in *CO* expression.

To assess the kinetics of the effect of light on *FT* mRNA levels, we analysed *FT* expression in seedlings grown for a week in short days and then transferred at dusk to light or darkness for 8 h. In wild-type seedlings, *FT* mRNA levels increased greatly during the first hours of extension of the illuminated part of the day (Fig. 5a, b). In contrast, this acute effect of light was notably reduced or completely absent in *cry2* and *phyA* mutants exposed to fluorescent (Fig. 5a) or incandescent (Fig. 5b) white light respectively. That *FT* expression was

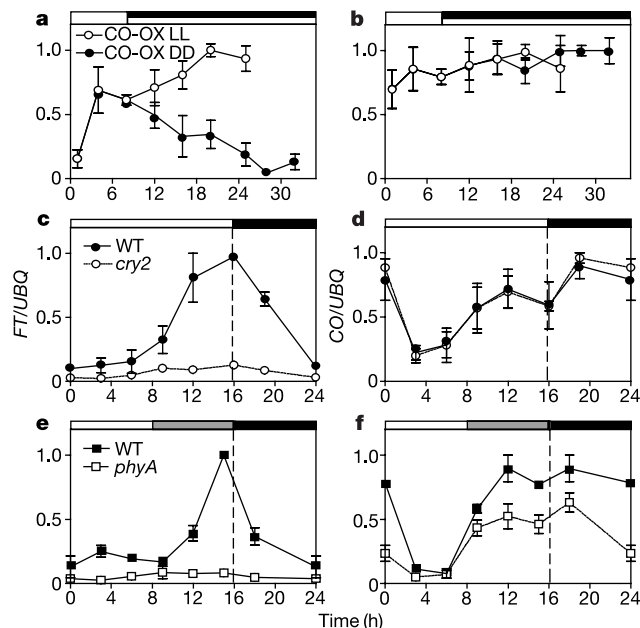


Figure 4 *FT* and *CO* expression in *CO*-overexpressing plants, and in wild-type, *cry2* and *phyA* mutants. **a, b**, *CO*-overexpressing plants (*CO*-OX) grown in short days (8 L: 16 D) and transferred to continuous light (LL) or darkness (DD) for 24 h. **c, d**, Wild-type seedlings and *cry2* mutants grown in long days (16 L: 8 D). **e, f**, Wild-type seedlings and *phyA* mutants grown under 8 h of fluorescent light extended with 8 h of incandescent light followed by 8 h of darkness. Data are mean $\pm$ s.e.m. of 2–3 independent experiments. White and grey bars represent light emitted from white-light fluorescent tubes or incandescent white-light bulbs respectively; dark bars represent dark periods.

upregulated more than 15-fold within the first 4 h of extension of a short-day photoperiod strongly suggests that the promotion of *FT* expression by *cry2* and *phyA* mainly results from a direct effect of light rather than from an indirect effect of photoperiod in clock entrainment. The effects of *cry2* and *phyA* on *FT* expression occurred through CO, because light induction of *FT* mRNA levels was greatly impaired in the *co-2* mutant (Fig. 5c, d), whereas no significant difference in *FT* mRNA levels was observed between *co-2* and wild-type plants kept in darkness at times when CO expression was close to peak levels in wild-type seedlings (Fig. 5c, d and Supplementary Information Fig. 3). The conclusion that *cry2* and *phyA* directly regulate *FT* expression by affecting CO function is consistent with the observation that far-red and blue light were the most effective wavelengths in maintaining high levels of *FT* mRNA in CO-overexpressing plants transferred to constant-light conditions (Supplementary Information Fig. 4). Light-dependent regulation of *FT* mRNA could result from effects of *cry2* and *phyA* on CO levels, on the intrinsic activity of the CO protein, or on the activity or levels of other molecule(s) that could be necessary for CO to regulate *FT* expression. Finally, we note that at this stage of development a single exposure to 8 h of light does not promote flowering²⁰. Therefore, the acceleration of flowering time by long days requires elevated *FT* mRNA levels over the course of several days and/or removal of an unknown factor that blocks *FT* activity.

For several decades, our understanding of how plants, insects, birds and mammals were able to measure daylength was based mainly on descriptive physiological experiments, which led to the development of two major models for photoperiodic time measurement. The external coincidence model states that photoperiodic responses are triggered by the coincidence between the illuminated part of the day and a photo-inducible phase of the circadian cycle²¹. In this model, light has two different actions: entraining the circadian system and inducing photoperiodic responses when present at the photo-inducible phase of the cycle²². In the internal coincidence model, a photoperiodic response is promoted by the internal coincidence of two or more circadian rhythms whose overlap is favoured by inductive photoperiods²³. In this model, light only acts by its effect on clock entrainment. It is not possible to

rule out that long days accelerate flowering time in wild-type seedlings in part by favouring an overlap in the timing of expression of some flowering-time genes. However, here we provide clear evidence that light, in addition to its well-established role in clock entrainment²⁴, affects flowering time through a direct effect on the expression of the flowering time gene *FT*. The induction of *FT* expression by light requires CO, and we demonstrate that exquisitely regulated clock control of the timing of CO expression is essential for photoperiodic discrimination. Therefore, our work firmly establishes that daylength regulation of flowering time in *Arabidopsis* is based, at least in part, on the coincidence of light, as perceived by *cry2* and *phyA*, with a particular circadian phase characterized by high levels of CO expression. This external coincidence produces elevated levels of *FT* mRNA, which finally promote the transition from vegetative to reproductive development (Supplementary Information Fig. 5). □

Methods

Plant material and growth conditions

The ecotype Landsberg *erecta* was used for all the experiments described here. The *toc1-1* mutant, originally isolated in the C-24 ecotype, was backcrossed five times to the wild type of the ecotype Landsberg *erecta*. The *cry2* (*fla-1* allele), *phyA-201* and *co-2* mutants were obtained from ABRC. The *toc1-1 co-2* and the *toc1-1 ft-2* double mutants were selected from segregating F₂ populations by polymerase chain reaction (PCR) with the primers *toc1-1f* GCGCCAGCTATCCACATTCC and *toc1-1r* TCAAGTTCCCAAGCATCATC, *co-2f* CCACGTGTAGGCACCTCAGGA and *co-2r* GAACAGCCACGAAGCAACCT, and *ft-2f* CCTGCTACAACCTGGAACAACC and *ft-2r* ATAGGCATCATCACCCTTCG, followed by digestion with DdeI, PciI or NlaIV to detect the *toc1-1*, the *co-2* or the *ft-2* alleles respectively. Seeds from a line overexpressing the CO gene from the 35S promoter were provided by G. Coupland (John Innes Center, Norwich, UK). Plants were grown for 8 days on agar medium under short days of 24 h (8 L: 16 D), 21 h (7 L: 14 D) or 30 h (10 L: 20 D), or under long days of 24 h (16 L: 8 D) and 21 h (14 L: 7 D). Light was provided by fluorescent white-light tubes (180 and 90 $\mu\text{mol m}^{-2} \text{s}^{-1}$ in short and long days, respectively) unless indicated otherwise.

Analysis of gene expression

Total RNA was extracted using RNeasy (Qiagen) according to the manufacturer's instructions. CO, *FT* and *UBQ* genes were quantified by PCR with reverse transcription of RNA (RT-PCR) as previously described^{1,25}. Gene-expression data were represented relative to the maximum value among all data sets in each figure (irrespective of genotype and environmental condition), after normalization to the *UBQ* control.

Analysis of flowering time

The plants were grown in soil under short days of 24 h (8 L: 16 D) or 30 h (10 L: 20 D). Time to flowering was taken as the number of rosette leaves at time of a 1-cm-high flower bolt.

Received 18 March; accepted 2 July 2002; doi:10.1038/nature00996.

- Putterill, J., Robson, E., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
- Samach, A. et al. Distinct role of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
- Kardailsky, I. et al. Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
- Onouchi, H., Igeno, M. I., Perilleux, C., Graves, K. & Coupland, G. Mutagenesis of plants overexpressing *CONSTANS* demonstrates novel interactions among *Arabidopsis* flowering-time genes. *Plant Cell* **12**, 885–900 (2000).
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
- Samach, A. & Coupland, G. Time measurement and the control of flowering in plants. *Bioessays* **22**, 38–47 (2000).
- Suarez-Lopez, P. et al. *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**, 1116–1120 (2001).
- Schaffer, R. et al. The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**, 1219–1229 (1998).
- Hicks, K. A. et al. Conditional circadian dysfunction of the *Arabidopsis* *early-flowering 3* mutant. *Science* **274**, 790–792 (1996).
- Park, D. et al. Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* *GIGANTEA* gene. *Science* **285**, 1579–1582 (1999).
- Zagotta, M. T. et al. The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering. *Plant J.* **10**, 691–702 (1996).
- Araki, T. & Komeda, Y. Analysis of the role of the late-flowering locus, *GI*, in the flowering of *Arabidopsis thaliana*. *Plant J.* **3**, 231–239 (1993).
- Huq, E., Tepperman, J. M. & Quail, P. H. *GIGANTEA* is a nuclear protein involved in phytochrome signaling in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **97**, 9789–9794 (2000).
- Strayer, C. et al. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* **289**, 768–771 (2000).
- Somers, D. E., Webb, A. A. R., Pearson, M. & Kay, S. The short-period mutant, *toc1-1*, alters circadian

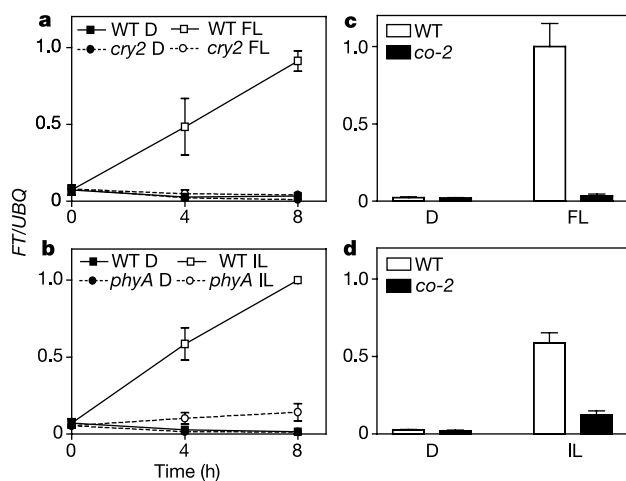


Figure 5 Acute induction of *FT* expression by light. The plants were grown for 7 days under cycles of 8 L: 16 D. At the end of the light period of the 8th day (indicated as time 0), the plants were exposed to an 8-h extension of either fluorescent (FL) or incandescent (IL) white light, or were kept in darkness (D) as controls. **a**, Wild-type (squares) and *cry2* mutant plants (circles) exposed to fluorescent light or transferred to darkness. **b**, Wild-type (squares) and *phyA* mutant plants (circles) exposed to incandescent light or transferred to darkness. **c**, Wild-type and *co-2* mutant plants exposed to fluorescent light or kept in darkness. **d**, Wild-type and *co-2* mutant plants exposed to incandescent light or kept in darkness. Data are mean $\pm$ s.e.m. of two independent experiments.

- clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. *Development* **125**, 485–494 (1998).
16. Millar, A. J., Carré, I. A., Strayer, C. A., Chua, N.-H. & Kay, S. A. Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163 (1995).
 17. Alabadi, D. et al. Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
 18. Guo, H. W., Yang, W. Y., Mockler, T. C. & Lin, C. T. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279**, 1360–1363 (1998).
 19. Johnson, E., Bradley, M., Harberd, N. P. & Whitelam, G. C. Photoresponses of light-grown *phyA* mutants of *Arabidopsis*. Phytochrome A is required for the perception of daylength extensions. *Plant Physiol.* **105**, 141–149 (1994).
 20. Corbesier, L., Gadiette, L., Silvestre, G., Jacqumard, A. & Bernier, G. Design in *Arabidopsis* of a synchronous system of floral induction by one long day. *Plant J.* **9**, 947–952 (1996).
 21. Bünning, E. Die endogene Tagesrhythmik als Grundlage der photoperiodischen Reaktion. *Ber. Dtsch Bot. Ges.* **54**, 590–607 (1936).
 22. Pittendrigh, C. S. & Minis, D. H. The entrainment of circadian oscillations by light and their role as photoperiodic clocks. *Am. Nat.* **98**, 261–294 (1964).
 23. Pittendrigh, C. S. Circadian rhythms and the circadian organization of living systems. *Cold Spring Harbor Symp. Quant. Biol.* **25**, 159–184 (1960).
 24. Yanovsky, M. J. & Kay, S. A. Signaling networks in the plant circadian system. *Curr. Opin. Plant Biol.* **4**, 429–435 (2001).
 25. Blazquez, M. A. & Weigel, D. Independent regulation of flowering by phytochrome B and gibberellins in *Arabidopsis*. *Plant Physiol.* **120**, 1025–1032 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. J. Casal, S. Harmer, P. Mas and F. Harmon for critical reading of the manuscript. This work was supported by an NIH grant to S.A.K. The work of M.J.Y. was initially supported by Conicet, Antorchas and the University of Buenos Aires and, more recently, by the Pew Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A.K. (e-mail: stevek@scripps.edu).

A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*

Liaoteng Wang*, Christian R. Eckmann†, Lisa C. Kadyk†‡, Marvin Wickens* & Judith Kimble*†

* Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

† Howard Hughes Medical Institute, 433 Babcock Drive, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

Messenger RNA regulation is a critical mode of controlling gene expression. Regulation of mRNA stability and translation is linked to controls of poly(A) tail length^{1,2}. Poly(A) lengthening can stabilize and translationally activate mRNAs, whereas poly(A) removal can trigger degradation and translational repression. Germline granules (for example, polar granules in flies, P granules in worms) are ribonucleoprotein particles implicated in translational control³. Here we report that the *Caenorhabditis elegans* gene *gld-2*, a regulator of mitosis/meiosis decision and other germline events⁴, encodes the catalytic moiety of a cytoplasmic poly(A) polymerase (PAP) that is associated with P granules in early embryos. Importantly, the GLD-2 protein sequence has diverged substantially from that of conventional eukaryotic PAPs, and lacks a recognizable RRM (RNA recog-

nition motif)-like domain. GLD-2 has little PAP activity on its own, but is stimulated *in vitro* by GLD-3. GLD-3 is also a developmental regulator, and belongs to the Bicaudal-C family of RNA binding proteins⁵. We suggest that GLD-2 is the prototype for a class of regulatory cytoplasmic PAPs that are recruited to specific mRNAs by a binding partner, thereby targeting those mRNAs for polyadenylation and increased expression.

We cloned the *gld-2* gene and analysed its transcripts (Fig. 1). The *gld-2* genomic region was identified by mutant rescue and RNA-aided interference (RNAi; see Methods) as well as elucidation of the molecular lesions in two *gld-2* mutants (see below). The *gld-2* gene encodes multiple mRNAs (Fig. 1a, b). A 5' probe detected a 4.7-kilobase (kb) band in wild-type poly(A)⁺ RNAs, but not in RNA from germline-less mutants (Fig. 1b, left). Therefore, this 4.7-kb mRNA appears to be germline-specific. Middle and 3' probes detected two somatic *gld-2* RNAs, of 4.6 and 4.0 kb (Fig. 1b, middle and right). These smaller *gld-2* mRNAs harbour distinct 5' terminal exons spliced to common exons (Fig. 1a). Two *gld-2* mutations identified genetically⁴ carried lesions in common exons: a predicted null mutant, *gld-2(q497)*, is a premature nonsense codon, and *gld-2(h292)* is a missense mutation (E875K) (Fig. 1a).

Because of our interest in *gld-2* germline functions, we focused on its 4.7-kb mRNA. Northern analysis (Fig. 1b, left) showed that this mRNA was abundant in embryos, fourth larval stages (L4s) and adults (Fig. 1c); *in situ* hybridization showed that it was abundant in the meiotic pachytene region and in oogenesis (Fig. 1d, e), but decreased during spermatogenesis (Fig. 1e). We did not detect the mRNA in putative null mutant *gld-2(q497)* (Fig. 1f), or with a sense-strand probe (anti-5') (Fig. 1g). Therefore, *gld-2* is expressed

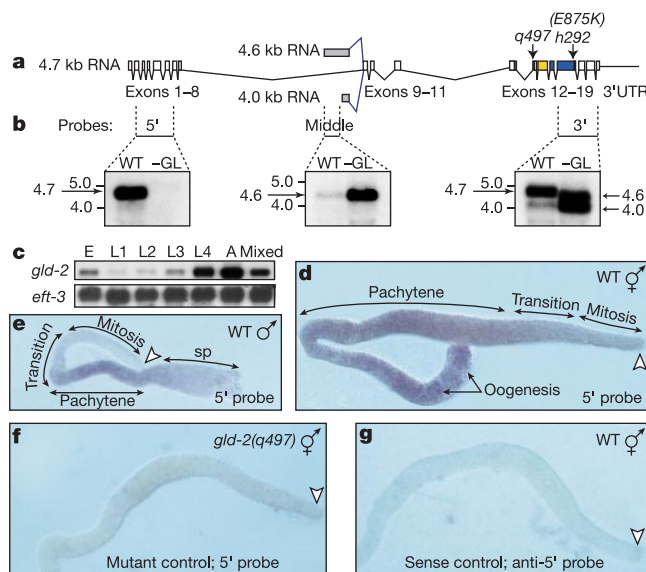


Figure 1 The *gld-2* gene and its transcripts. **a**, *gld-2* exon/intron structure. Exons, open boxes; introns, thin lines. Colour coding as in Fig. 2b. **b**, Top, probes (see Methods). Bottom, northern blots of poly(A)⁺ RNAs from mixed stage wild-type (WT) animals or from *gld-1* mutant adults with no germ line (—GL). Size markers in kb. Arrows, *gld-2* transcripts. Sizes of *gld-2* transcripts on northern blots (4.7 kb, 4.6 kb and 4.0 kb) correspond to sizes predicted by cDNA analyses (4,533 nt, 4,273 nt and 3,691 nt, excluding the poly(A) tail). **c**, Developmental expression of *gld-2* mRNA. Northern blot of poly(A)⁺ RNAs from staged animals. E, embryo; L1–L4, first–fourth larval stage; A, adult; mixed, mixed stages. Above, 5' probe, Fig. 1b; below, loading control. **d–g**, *In situ* hybridization of dissected germ lines. **d–f**, 5' probe, Fig. 1b; open arrowhead, distal end of germ line. **d**, Germ line of wild-type hermaphrodite adult. **e**, Germ line of wild-type male adult; sp, spermatogenesis; WT, wild type; **f**, germ line of *gld-2(q497)* homozygous mutant adult; **g**, germ line of wild-type hermaphrodite adult, probed with sense strand of cDNA fragment covering exons 2–8 (5' probe, 1b).

‡ Present address: Exelixis, Inc., 170 Harbor Way, PO Box 511, So. San Francisco, California 94083-0511, USA.

in the germ line and is developmentally regulated.

Database searches revealed that GLD-2 protein belongs to the DNA polymerase β -like superfamily of nucleotidyltransferases (NT) (Fig. 2a; refs 6, 7). Specifically, GLD-2 is a group 2 NT member, including DNA polymerase σ of *Saccharomyces cerevisiae* (also known as pol κ and Trf4p) and eukaryotic PAPs (Fig. 2a). GLD-2 architecture and sequence is divergent from that of canonical PAPs (Fig. 2b, d), but similar to a different cluster of NT family members (Fig. 2e). GLD-2 contains three critical carboxylate side chains essential for catalytic activity (Fig. 2c, red) present in all DNA polymerase β superfamily members; furthermore, GLD-2 possesses putative ATP-interacting residues (Fig. 2c, green; Fig. 2d, green). Classical PAPs have a catalytic region (Fig. 2c, gold), a 'central' domain (Fig. 2c, blue), and an RRM-like region (Fig. 2c, violet)^{8,9}. By sequence comparison, GLD-2 harbours catalytic and central domains (Fig. 2b, d, colour-coded overlines), but is highly diverged from classical eukaryotic PAPs, including *C. elegans* PAP-1 (C. Luitjens and M.W., unpublished results) (Fig. 2d). Classical PAPs show extensive amino-acid conservation among themselves, but limited conservation with GLD-2 (Fig. 2d, black and grey boxes). Outside its catalytic and central domains, GLD-2 shares little similarity to canonical PAPs; in particular, GLD-2 has no apparent RRM-like region (Fig. 2b), which is thought to be critical for PAP RNA binding^{8,9}. Therefore, GLD-2 shares some key features with classical PAPs, but is divergent in motif architecture and amino acid sequence.

To examine GLD-2 protein, we generated polyclonal antibodies to the amino-terminal region (Fig. 2b) and detected a prominent protein of relative molecular mass 125,000 (M_r 125K) on western blots (Fig. 3a, lanes 1, 4, 5). This protein, which corresponds in size to the predicted product of the germline *gld-2* mRNA, was detected in *gld-2(h292)* homozygotes and *gld-2(q497)/+* heterozygotes (Fig. 3a, lanes 6, 7), but not in *gld-2(q497)* homozygotes (Fig. 3a, lane 8). Pre-immune serum did not recognize this band, but detected others that served as a loading control (not shown). We conclude that the α -GLD-2 antibody recognizes GLD-2, that the *gld-2(h292)* mutant produces a nearly wild-type level of protein and that *gld-2(q497)* is a strong loss-of-function or null allele.

By immunocytochemistry, GLD-2 was found to be predominantly cytoplasmic in both germ line (Fig. 3b) and early embryo (Fig. 3c). Within the germ line, GLD-2 was detectable in the mitotic region and became abundant during pachytene and oogenesis (Fig. 3b). GLD-2 decreased during spermatogenesis in both sexes, and was undetectable in mature sperm (not shown). In early embryos, GLD-2 was diffuse in the cytoplasm of early P0 embryos, co-localized with P granules in late P0 embryos and remained associated with P granules in germline blastomeres (Fig. 3c, not shown). P granules are essential for germline development^{3,10}. In ~100-cell embryos, GLD-2 was undetectable.

Given its presence in oocytes and early embryos, we tested whether GLD-2 was required for embryogenesis. To deplete both maternal and zygotic *gld-2* mRNAs, wild-type adult hermaphrodites were treated with double-stranded RNA corresponding to either the *gld-2* germline-specific region (exons 2–8) or its common region (exons 16–18) to produce *gld-2(RNAi)* embryos (see Methods). In both cases, most *gld-2(RNAi)* embryos failed to hatch (99%, $n > 500$ in 26–36 h period after treatment). To visualize chromosomes in *gld-2(RNAi)* embryos, we used a strain carrying a histone:GFP transgene (AZ212)¹¹. Whereas mock-treated AZ212 embryos cleaved normally (Fig. 3d), *gld-2(RNAi)* AZ212 embryos did not cleave and possessed malformed nuclei in clusters (Fig. 3e). We conclude that *gld-2* activity is required for embryogenesis, and that GLD-2 protein co-localizes with P granules.

A specific interaction between GLD-2 and another germline regulator, GLD-3 (ref. 5), was discovered in yeast two-hybrid screens. Specifically, using GLD-2 as 'bait', 2,000,000 transformants

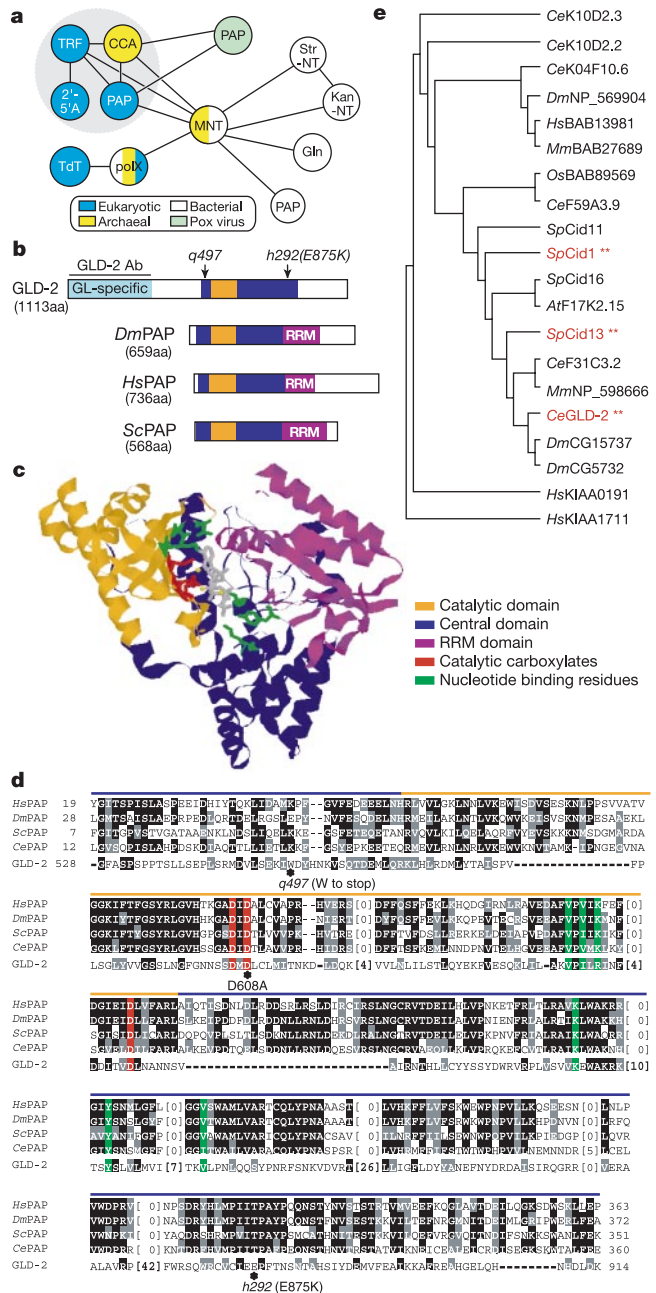


Figure 2 GLD-2 belongs to the polymerase β nucleotidyltransferase superfamily. **a**, The polymerase β superfamily (adapted from ref. 7). Small colour-coded circles, families; large grey circle, group 2 families. CCA, CCA-adding enzymes; 2'-5' A, 2'-5' oligoA synthetases; TRF, Trf4p-like proteins; other acronyms as in ref. 7. **b–d**, Colour coding based on crystal structures of bovine and yeast PAPs^{8,9}. Gold, catalytic domain; blue, central domain; violet, RRM domain. **b**, GLD-2 and PAP domains compared. *Drosophila* (*Dm*), human (*Hs*) and yeast (*Sc*). GLD-2 domains identified by Pfam search²⁷. aa, amino acid. **c**, Bovine PAP 3D structure, with key residues shown in stick form (adapted from ref. 8). Created by RasMol based on PDB file 1F5A (for bovine PAP). **d**, Amino-acid sequence alignment of GLD-2 and PAP core regions based on clustalW output²⁸ and polymerase β superfamily analyses²⁹. Mutants designated below. Red, catalytic residues; green, required for ATP binding. **e**, Unrooted tree of GLD-2 and its homologues, created with PHYLIP programs³⁰, based on ClustalW alignment using parsimony. Species are: *Ce*, *C. elegans*; *Dm*, *Drosophila*; *Hs*, human; *Mm*, mouse; *Os*, rice; *Sp*, *S. pombe*; *At*, *Arabidopsis*. Only homologues with E -values less than 1×10^{-10} in the first PSI blast were used; tree was built using the catalytic and central domain sequences as in **d** (GLD-2 amino acids 528–914 and corresponding sequences of its homologues). Cid1 and GLD-2 (shown red) both function in cell cycle control; Cid13 (shown red) is involved in the replication stress response²²; functions of others are unknown.

were screened and 30 *gld-3* cDNAs (T07F8.3) found; using GLD-3 as bait, 1,500,000 transformants were screened and 94 *gld-2* cDNAs recovered. To identify the region of GLD-2 critical for GLD-3 binding, GLD-2 variants were assayed for GLD-3 interaction. A GLD-2 fragment comprising both catalytic and central domains was essential (amino acids 544–924) (Fig. 4a). A GLD-2-E875K mutant, designed after *gld-2(h292)*⁴, interacted poorly with GLD-3 (Fig. 4a, E875K and Δ7). Indeed, β-galactosidase activity was reduced 7- to 16-fold by GLD-2(h292)-E875K (Fig. 4a, compare for example Δ2 to Δ7), but GLD-2 levels were equivalent (Fig. 4b). Importantly, GLD-2-E875K was present at normal levels in *C. elegans* (Fig. 3a, lane 6), even though it disrupts *gld-2* function. We conclude that GLD-2 binds specifically to GLD-3, and that GLD-2-E875K is defective in GLD-3 binding. Therefore, the GLD-2/GLD-3 interaction appears to be important for development.

Given its sequence similarity to nucleotidyltransferases and its cytoplasmic location, we considered that GLD-2 might be a cytoplasmic PAP, even though its architecture and sequence diverged substantially from classical PAPs. To test this idea, we initially assayed incorporation of radiolabelled ATP into an RNA substrate. Specifically, GLD-2 was translated *in vitro*, either on its own or together with GLD-3. The *in vitro* translation mixture was incubated with ³²P-ATP and an unlabelled poly(A) substrate, and incorporation of label into acid-insoluble material was measured (see Methods). GLD-2 on its own had low activity, whereas GLD-3 had none; however, GLD-2 and GLD-3 together gave a robust response (Fig. 4c). We also measured incorporation in three control reactions (no protein and two GLD-2 mutants together with GLD-3). GLD-2-D608A was designed to abolish the catalytic site (Fig. 2c) and GLD-2-E875K was used to disrupt GLD-3 binding (Fig. 4a). The control reactions yielded no measurable ³²P-ATP incorporation (Fig. 4c). From these experiments, we argue that GLD-2 is in fact a nucleotidyltransferase and that both its predicted active site and GLD-3 binding region are essential for enzymatic activity.

We next analysed the products of the GLD-2/GLD-3 nucleotidyltransferase activity by electrophoresis and autoradiography (Fig. 4d). To this end, reactions were done as described above, except that C₃₅A₁₀ (see Methods) was used as substrate. Two

exposures of the same autoradiogram are shown (Fig. 4d). As a marker, C₃₅A₁₀ was 3' end-labelled with cordycepin triphosphate ([α-³²P] 3' dATP) (C₃₅A₁₀*dA; Fig. 4d left, lane 1). GLD-2 by itself exhibited modest incorporation from ATP into bands that extended the substrate by only one or a few nucleotides (Fig. 4d left, lane 2). In contrast, GLD-2 plus GLD-3 stimulated incorporation, resulting in more product with a 'ladder' of poly(A) extending the substrate more than 30 adenosines (Fig. 4d, lane 4). The ladder mimics the activity of bovine nuclear PAP (bPAP), but is less efficient (Fig. 4d, compare lanes 4 and 7). This difference may reflect the fact that bovine PAP acts as a monomer, whereas GLD-2 PAP activity is dependent on the interaction of two dilute proteins. Furthermore, although abundant products had only two or three nucleotides added (asterisks in Fig. 4d, lane 4), more-minor products had as many as 70 additional nucleotides. We conclude that GLD-2/GLD-3 can catalyse the addition of a poly(A) tail to an RNA substrate.

Four controls support the conclusion that GLD-2 is a PAP. First, GLD-2 PAP activity was abolished by a site-directed mutation in the inferred active site (D608A) (Fig. 4d, lane 5). Importantly, GLD-2-D608A level is equivalent to that of wild-type GLD-2 in the same assay (Fig. 4d SDS–polyacrylamide gel electrophoresis, SDS–PAGE, compare lanes 4 and 5). Thus, the GLD-2 putative active site is required for AMP addition *in vitro*. Second, GLD-2 PAP activity was abolished by the E875K mutation (Fig. 4d, lane 6), which disrupts GLD-2/GLD-3 binding (Fig. 4a). The GLD-2-E875K level was equivalent to wild-type GLD-2 (Fig. 4d, compare lanes 4 and 6). Third, GLD-2-dependent incorporation is substrate dependent and requires ATP (not shown). Thus, replacement of ATP with GTP, CTP or UTP did not yield incorporation onto the substrate. Finally, products produced by GLD-2 plus GLD-3 were selectively retained on oligo(dT) cellulose, suggesting they were polyadenylated (not shown).

The GLD-2/GLD-3 enzyme represents a new type of poly(A) polymerase (Fig. 5). Canonical PAPs, which include nuclear and cytoplasmic enzymes, are all closely related^{12–15}; they are monomeric and possess three key domains (Fig. 5, left)^{8,9}. By contrast, GLD-2 appears to function as a heterodimer (Fig. 5, right). GLD-2 harbours the catalytic and central domains; GLD-3 has five consecutive K homology (KH)-related motifs⁵ which may, at least in

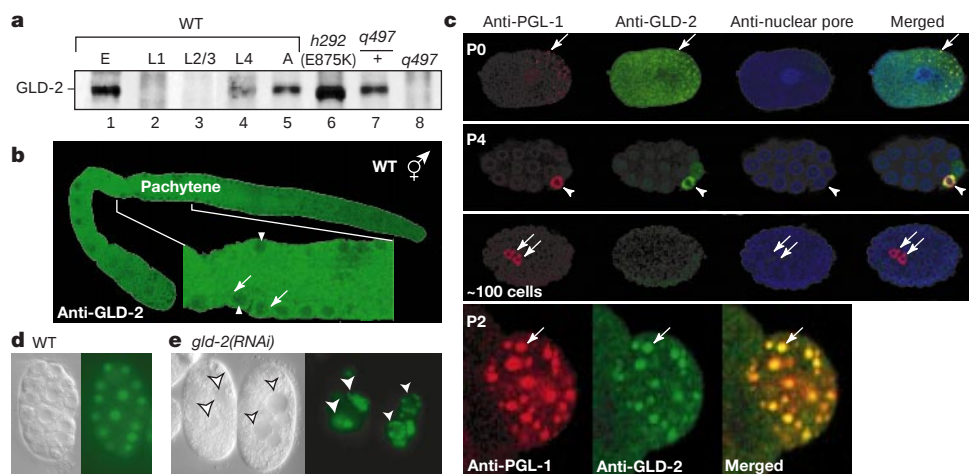


Figure 3 The GLD-2 protein. Polyclonal anti-GLD-2 antibodies were affinity purified. **a**, Western blot of proteins from wild-type embryos (E), larvae (L1–L4), and adults (A) (lanes 1–5), and adults of genotype *gld-2(h292)/gld-2(h292)* (lane 6), *gld-2(q497)/gld-2(+)* (lane 7), and *gld-2(q497)/gld-2(q497)* (lane 8). **b**, GLD-2 protein is in germline cytoplasm. Extruded WT adult hermaphrodite germline; GLD-2 is abundant in pachytene region and oocytes. Magnified view shows lack of GLD-2 in nuclei (arrowheads) and presence of GLD-2 in granular form (arrows). A control *gld-2(q497)* extruded germline showed no anti-GLD-2 staining (not shown). **c**, GLD-2 protein is associated with P granules in early

embryos. Embryos stained with antibody to P granule marker, PGL-1¹⁰, to GLD-2, and to nuclear pore antigen. Top, late P0 embryo, GLD-2 co-localizes with P granules; second panel down, 28-cell embryo, P4, white arrowhead; third panel, ~100-cell embryo, germline precursor cells, Z2 and Z3, arrows; bottom, magnified view of P2 blastomere to show PGL-1 and GLD-2 co-localization (arrows). **d, e**, Transgenic strain AZ212. Left, Nomarski image; right, nuclei visualized by histone::GFP marker. Both control and *gld-2(RNAi)* embryos are of approximately same age. **d**, Mock injected control. **e**, *gld-2(RNAi)*.

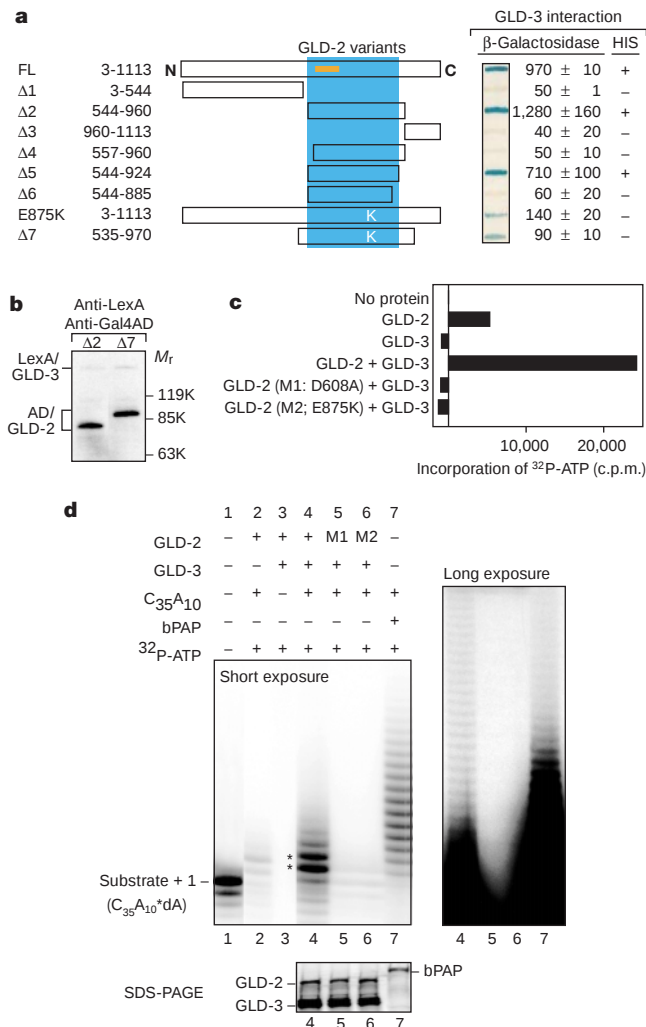


Figure 4 GLD-2/GLD-3 is a new type of poly(A) polymerase. **a**, Left, GLD-2 deletions used in yeast two-hybrid assays to map region of GLD-3 interaction. Right, results of both filter and liquid β -galactosidase assays as well as a growth assay (HIS). **b**, GLD-2 fragments were expressed at similar level. Western blot, $\Delta 2$ and $\Delta 7$ fragments as in **a**. **c**, Nucleotidyltransferase assay. Incorporation of ^{32}P -ATP was measured in reticulocyte lysates programmed with plasmids encoding GLD-2, GLD-3 or variants. Data reported as c.p.m., not molar quantities, because ATP concentration in lysate was not known. The lysate exhibits a background incorporation (10,000 c.p.m.) independent of GLD-encoding plasmids, which has been subtracted here. In the experiment shown, 1 mM MnCl_2 was added to the lysate; similar experiments with added MgCl_2 reduced incorporation fourfold. **d**, Poly(A) polymerase assay. Reaction products analysed on a 12% sequencing gel and visualized by autoradiography. Left, shorter exposure; right, longer exposure. Below, SDS-PAGE showing that proteins were expressed at similar levels. M1, D608A; M2, E875K; C₃₅A₁₀, substrate; bPAP, bovine poly(A) polymerase.

part, substitute for the RRM domain of classical PAPs. In the simplest view, GLD-2 and GLD-3 act together as a heterodimer to accomplish what classical PAPs do on their own. However, we suggest that GLD-2 is tailored for a more regulatory role than that typical of classical PAPs. For example, GLD-3 is likely to provide sequence specificity to the GLD-2 catalytic activity, and GLD-2 may interact with additional partners to expand its repertoire of regulation.

GLD-2 and GLD-3 are likely to function together during nematode development. First, GLD-2 and GLD-3 have similar, albeit not identical, functions in germline development and embryogenesis (refs 4, 5, and this work). Second, both are cytoplasmic and associated with P granules (ref. 5, this work), large complexes of

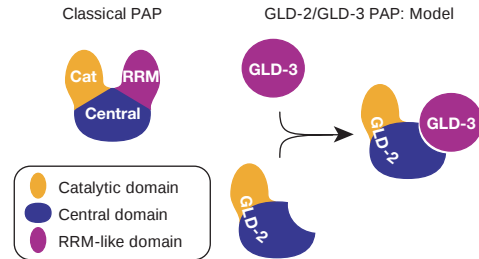


Figure 5 Model for architecture of GLD-2/GLD-3 rcPAP enzyme. Left, classical PAPs. Right, speculative architecture of GLD-2/GLD-3. Domains colour coded as in Fig. 2.

RNA and protein that are critical for germline development^{3,10}. GLD-2 and GLD-3 may polyadenylate mRNAs associated with P granules (for example, *nos-2*; ref. 16) or may be stored there for segregation to germline blastomeres. GLD-2 may be targeted to specific mRNAs by GLD-3, which is a Bic-C family KH protein⁵. Other KH proteins (FMRP, NOVA, hnRNP) bind RNAs through sequence-specific interactions^{17–20}. GLD-2 may also be targeted to specific mRNAs indirectly via the interaction of GLD-3 with FBF⁵. FBF is a sequence-specific RNA-binding protein and member of the PUF family²¹. PUF proteins appear to repress mRNAs by promoting poly(A) removal²¹. GLD-3 antagonizes FBF⁵, and works with GLD-2 to promote poly(A) addition (this work). Therefore, GLD-2/GLD-3 may switch FBF from a repressive to an activating mode.

Regulatory cytoplasmic PAPs of the GLD-2/GLD-3 class may be common. Within the large superfamily of DNA polymerase β -like nucleotidyltransferases, several are closely related to GLD-2 (Fig. 2e). To date, most have no assigned function, but *Schizosaccharomyces pombe* Cid13 and Cid1 appear to be rcPAPs^{22,31}. The similarity between GLD-2 and Cid1 is particularly striking, as both are involved in cell cycle control. GLD-2 promotes entry into meiosis at the expense of mitosis⁴, and Cid1 inhibits mitosis²³. We suggest that GLD-2 and Cid1 may in fact be components of an ancient regulatory circuit controlling the cell cycle, and that other GLD-2 relatives may similarly be regulatory cytoplasmic PAPs. □

Methods

Molecular cloning of *gld-2*

Three-factor mapping places *gld-2* 0.05 map unit to the right of *bli-4*. Cosmids in this region were injected into strain JK1716 [*bli-4(e937) gld-2(q497)/dpy-5(e61) unc-13(e51)*] or strain JK1732 [*bli-4(e937) gld-2(h292)/dpy-5(e61) unc-13(e51)*]. Cosmid ZC308 gave ~4% germline rescue.

Transcript analyses

Northern blots were performed as described²⁴. Templates for making RNA probes (*gld-2* 5', middle, 3'; *eft-3*) were made by polymerase chain reactions (PCRs) from pJK830, pJK831, pJK832 and pBluescript-*eft-3* (gift from P. Anderson). To determine the *gld-2* 3' end, semi-nested PCR was performed using $\Delta\text{AE.1}$, a *C. elegans* mixed-stage oligo(dT) primed complementary DNA library (gift from A. Puoti). One PCR product was confirmed and sequenced. A stretch of 22 As was found at the end of the 3' untranslated region (UTR). To determine the *gld-2* 5' ends, reverse transcriptions (RT) were performed using SuperScript II Reverse Transcriptase (Gibco BRL) and poly(A)⁺ RNA from either wild-type mixed-stage worms or *gld-1(q224)* mutants raised at 25 °C, which have no germ line. The resultant cDNAs were then used as templates for semi-nested PCR with SL1 (a trans-spliced leader in *C. elegans*) as the constant 5' primer. All PCR products were cloned into pSTBlue-1 and sequenced. The 4.7-kb mRNA is SL1 trans-spliced, comprises 19 exons including an 86-nucleotide 5' UTR and 1,105-nucleotide 3' UTR.

Antibody production, western blot and immunocytochemistry

Polyclonal antibodies were generated from rabbits using a keyhole limpet haemocyanin (KLH)-conjugated peptide corresponding to GLD-2 amino acids 108–127 (Genemed Synthesis) or from rats using a GST-GLD-2 fusion protein carrying amino acids 13–330 of GLD-2. Rabbit anti-PGL-1 antibody was a gift from S. Strome. Monoclonal antibody 414, the anti-nuclear pore monoclonal, was purchased from BABCO. Western blots were performed using the GLD-2 peptide antibody as described²⁴. Immunocytochemistry followed published procedures²⁵ using the GST-GLD-2 fusion-protein antibody, which was specific for GLD-2 as demonstrated on *gld-2(q497)* extruded germ lines and *gld-2(RNAi)* embryos.

RNAi

Double-stranded RNAs (dsRNAs) were made using *gld-2* cDNAs (pJK830, exons 2–8 or pJK831, exons 16–18) as templates. Young adults were either injected with $2 \mu\text{g} \mu\text{l}^{-1}$ *gld-2* dsRNA or soaked in $10 \mu\text{l}$ of $2 \mu\text{g} \mu\text{l}^{-1}$ *gld-2* dsRNA for 12 h at 20°C or mock-treated by injection with M9 buffer. Embryos were collected at defined intervals after treatment and processed together.

Poly(A) polymerase assay

Proteins were *in vitro* translated using the TNT coupled transcription–translation system (Promega), and assayed using buffer conditions essentially as described²⁶. For scintillation counting, poly(A) (Roche) was used as substrate. For gel assays, we used RNA oligo, C₃₅A₁₀ (Dharmacon), a 45-nucleotide and supplemental 1 mM MgCl₂. Products were analysed on 12% sequencing gels.

Received 8 May; accepted 16 July 2002; doi:10.1038/nature01039.

- Richter, J. D. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 785–805 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Wickens, M., Goodwin, E. B., Kimble, J., Strickland, S. & Hentze, M. W. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 295–370 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Seydoux, G. & Strome, S. Launching the germline in *Caenorhabditis elegans*: regulation of gene expression in early germ cells. *Development* **126**, 3275–3283 (1999).
- Kadyk, L. C. & Kimble, J. Genetic regulation of entry into meiosis in *Caenorhabditis elegans*. *Development* **125**, 1803–1813 (1998).
- Eckmann, C., Kraemer, B., Wickens, M. & Kimble, J. GLD-3, a Bicaudal-C homolog that represses FBF to control germline sex determination in *C. elegans*. *Dev. Cell* (in the press).
- Holm, L. & Sander, C. DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem. Sci.* **20**, 345–347 (1995).
- Aravind, L. & Koonin, E. V. DNA polymerase β -like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res.* **27**, 1609–1618 (1999).
- Martin, G., Keller, W. & Double, W. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J.* **19**, 4193–4203 (2000).
- Bard, J. *et al.* Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science* **289**, 1346–1349 (2000).
- Kawasaki, I. *et al.* PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in *C. elegans*. *Cell* **94**, 635–645 (1998).
- Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**, 1217–1226 (2001).
- Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755–2766 (1997).
- Kashiwabara, S.-i. *et al.* Identification of a novel isoform of poly(A) polymerase, TPAP, specifically present in the cytoplasm of spermatogenic cells. *Dev. Biol.* **228**, 106–115 (2000).
- Kyriakopoulou, C. B., Nordvang, H. & Virtanen, A. A novel nuclear human poly(A) polymerase (PAP), PAP γ . *J. Biol. Chem.* **276**, 33504–33511 (2001).
- Topalian, S. L. *et al.* Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. *Mol. Cell. Biol.* **21**, 5614–5623 (2001).
- Subramaniam, K. & Seydoux, G. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* **126**, 4861–4871 (1999).
- Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K. & Darnell, R. B. The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl Acad. Sci. USA* **97**, 5740–5745 (2000).
- Brown, V. *et al.* Microarray identification of EMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487 (2001).
- Darnell, J. C. *et al.* Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489–499 (2001).
- Ostarek, D. H. *et al.* mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* **89**, 597–606 (1997).
- Wickens, M., Bernstein, D. S., Kimble, J. & Parker, R. A PUF family portrait: 3' UTR regulation as a way of life. *Trends Genet.* **18**, 150–157 (2002).
- Saitoh, S. *et al.* Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell* **109**, 563–573 (2002).
- Wang, S. W., Toda, T., MacCallum, R., Harris, A. L. & Norbury, C. Cid1, a fission yeast protein required for S-M checkpoint control when DNA polymerase delta or epsilon is inactivated. *Mol. Cell. Biol.* **20**, 3234–3244 (2000).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (ed.) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Crittenden, S. L. & Kimble, J. in *Cell: A Laboratory Manual* (eds Spector, D., Goldman, R. & Leinwand, L.) 108.1–108.9 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1998).
- Lingner, J., Radtke, I., Wahle, E. & Keller, W. Purification and characterization of poly(A) polymerase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **266**, 8741–8746 (1991).
- Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313**, 903–919 (2001).
- Felsenstein, J. *PHYLIP (Phylogeny Inference Package) Version 3.5c* (Department of Genetics, Univ. Washington, Seattle, 1993).
- Read, R. L., Martinho, R. G., Wang, S.-W., Carr, A. M. & Norbury, C. J. Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. *Proc. Natl Acad. Sci. USA* (in the press).

Acknowledgements

We thank R. Read and C. Norbury for sharing unpublished observations, and S. Crittenden for comments on the manuscript. C.E. was supported by the Human Frontier Science Program, L.K. was supported by the American Cancer Society, J.K. is an investigator with the Howard Hughes Medical Institute, and M.W. is supported by the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.K. (e-mail: jekimble@facstaff.wisc.edu). The GenBank accession number of the GLD-2 cDNA sequence is AY125085.

Forkhead transcription factor FOXO3a protects quiescent cells from oxidative stress

Geert J. P. L. Kops[†], Tobias B. Dansen[‡], Paulien E. Polderman^{*}, Ingrid Saarloos^{*}, Karel W. A. Wirtz[‡], Paul J. Coffey[§], Ting-T. Huang^{||}, Johannes L. Bos^{*}, René H. Medema[¶] & Boudewijn M. T. Burgering^{*#}

^{*} Department of Physiological Chemistry, University Medical Center Utrecht and Center for Biomedical Genetics, 3584 CG Utrecht, The Netherlands

[‡] Department of Biochemistry of Lipids, Institute of Biomembranes, Utrecht University, 3584 CH Utrecht, The Netherlands

[§] Department of Pulmonary Diseases, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands

^{||} Department of Pediatrics, University of California, San Francisco, California 94143, USA

[¶] Division of Molecular Biology, H8, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

[#] These authors contributed equally to this work

Reactive oxygen species are required for cell proliferation but can also induce apoptosis¹. In proliferating cells this paradox is solved by the activation of protein kinase B (PKB; also known as c-Akt), which protects cells from apoptosis². By contrast, it is unknown how quiescent cells that lack PKB activity are protected against cell death induced by reactive oxygen species. Here we show that the PKB-regulated Forkhead transcription factor FOXO3a (also known as FKHR-L1) protects quiescent cells from oxidative stress by directly increasing their quantities of manganese superoxide dismutase (MnSOD) messenger RNA and protein. This increase in protection from reactive oxygen species antagonizes apoptosis caused by glucose deprivation. In quiescent cells that lack the protective mechanism of PKB-mediated signalling, an alternative mechanism is induced as a consequence of PKB inactivity. This mechanism entails the activation of Forkhead transcription factors, the transcriptional activation of MnSOD and the subsequent reduction of reactive oxygen species. Increased resistance to oxidative stress is associated with longevity. The model of Forkhead involvement in regulating longevity stems from genetic analysis in *Caenorhabditis elegans*^{3–6}, and we conclude that this model also extends to mammalian systems.

Reactive oxygen species (ROS) are a primary cause of cellular damage that leads to cell death¹. In proliferating cells, protection from cell death is mediated by activity of the phosphatidylinositol-3-OH kinase (PI(3)K)–PKB signalling pathway, which is dependent on the presence of glucose². In the absence of PI(3)K–PKB signalling, the FOXO subfamily of Forkhead transcription factors, con-

[†] Present address: Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, California 92093-0670, USA.

sisting of FOXO4 (also known as AFX), FOXO1 (FKHR) and FOXO3a (FKHR-L1), is activated⁷⁻⁹. In most cell types, this leads to cell-cycle arrest and quiescence but not apoptosis, despite the absence of PKB activity^{10,11}. Similarly, in *C. elegans* an absence of PKB signalling activates the FOXO homologue DAF-16 and causes dauer formation—a complex phenotype characterized by increased resistance to oxidative stress³⁻⁶.

We have established an inducible FOXO3a cell line through stable transfection of a conditionally active HA-FOXO3a-A3-ER fusion into DLD-1 human colon carcinoma cells¹¹. The HA-FOXO3a-A3-ER fusion protein is constitutively expressed in quantities comparable to the endogenous protein in clone DL23 (Supplementary Information Fig. 1), but remains inhibited unless presented with a modified ligand for the estrogen receptor (ER), 4-hydroxy-tamoxifen (4-OHT)¹². Treatment of the DL23 cells with 4-OHT for 24 h resulted in the specific activation of FOXO3a, a strong increase in the amounts of p27^{kip1} protein and a decrease in cell proliferation (Supplementary Information Fig. 1), and forced these cells into a state of quiescence¹¹. Notably, 4-OHT had no effect on the control DLD-1 cells (Supplementary Information Fig. 1).

To measure the effect of Forkhead activity on cellular ROS, we loaded the DL23 cell line, either untreated or treated with 4-OHT for 16 h, with the lipophilic C11-BODIPY^{581/591} probe, which can be used to quantify lipid oxidation in single living cells¹³. The initial rate of oxidation was 4 times lower, and the total oxidized fraction of the probe after a 50-min treatment with hydrogen peroxide (H₂O₂) was 2–3 times lower, in cells containing active Forkhead than in control cells (Fig. 1a, b). It should be noted that activation of PKB can also protect cells against various stresses, and that H₂O₂ treatment has been shown to induce PKB activation¹⁴. In DLD-1 and DL23 cells, however, we observed no PKB activation by H₂O₂ at concentrations used to induce radical formation (200 μ M; Supplementary Information Fig. 1). Because treatment of the DLD-1 control cell-line with 4-OHT had no effect, these results show that the specific activation of Forkheads increases cellular protection against ROS.

We examined whether Forkheads increase cellular antioxidant capacity through the regulation of antioxidant enzymes, notably SODs. Treating the DL23 cell line with 4-OHT for up to 24 h showed a gradual increase in the amount of the mitochondrial MnSOD protein but no change in protein quantities of the cytoplasmic copper/zinc SOD (CuZnSOD; Fig. 1c). The increase in MnSOD expression was not a secondary event caused by p27^{kip1}-induced cell-cycle arrest, because wild-type and p27^{kip1}-/- mouse embryo fibroblasts (MEFs) showed a similar increase in MnSOD protein on infection with a retrovirus expressing FOXO3a-A3 (Fig. 1d). Similar effects were seen when we used a retrovirus expressing HA-FOXO4.

We investigated whether the regulation of MnSOD protein expression is under the control of endogenous PI(3)K–PKB–Forkhead signalling. Serum deprivation of 3T3-L6 cells increased the amounts of MnSOD protein (Fig. 1e), and subsequent activation of PI(3)K–PKB and inactivation of FOXO3a (indicated by phosphorylation of Thr 32) by the addition of insulin (Fig. 1e) reduced the amounts of MnSOD (Fig. 1e). This reduction was abolished when cells were pretreated with the PI(3)K inhibitor LY294002. These effects were also observed with the DLD-1 cell line (Supplementary Information Fig. 1). To establish the involvement of endogenous FOXO in the control of MnSOD, we expressed dominant-negative mutants of FOXO3a in 293T cells¹⁵. Notably, we could show that the LY294002-induced increase in MnSOD protein was prevented by expression of these dominant-negative mutants, which indicates that endogenous FOXO regulates the expression of MnSOD (Fig. 1f).

To investigate whether MnSOD is transcriptionally regulated by Forkheads, we examined the expression of MnSOD in the DL23 cell line. The levels of MnSOD mRNA were increased between 8 and 16 h of 4-OHT treatment and a roughly tenfold increase was seen after 24 h of FOXO3a activation (Fig. 2a). Similarly, when 3T3-L6 cells were serum starved for 48 h, the levels of MnSOD mRNA increased 3.5-fold, but were decreased after insulin treatment to

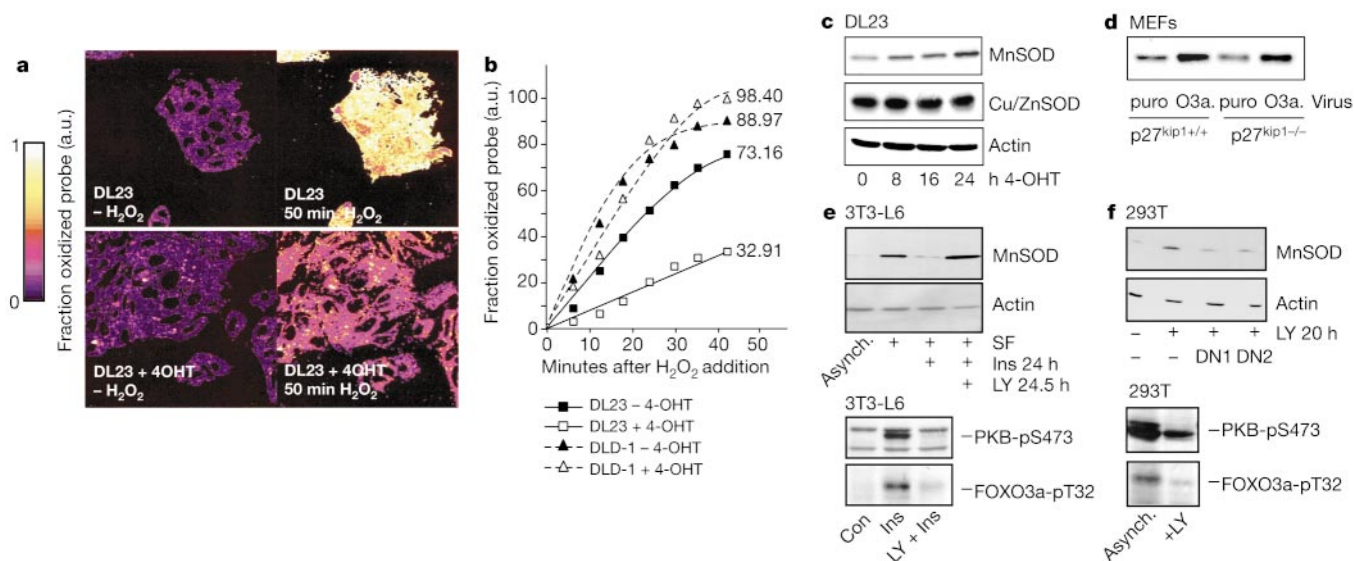


Figure 1 Forkhead transcription factors increase cellular protection against ROS. **a**, DL23 cells left untreated or treated with 500 nM 4-OHT for 16 h were loaded with the C11-BODIPY^{581/591} probe, and treated with 200 μ M H₂O₂ for 50 min. Images were taken by confocal microscopy¹³. Yellow areas represent high oxidation, purple areas represent low oxidation. **b**, DL23 (unbroken lines) and DLD-1 (broken lines) cells were treated as in **a**, and the fraction of oxidized probe per time point in arbitrary units (a.u.) was calculated. **c**, Total lysates of DL23 cells left untreated or treated with 500 nM 4-OHT were analysed for expression of MnSOD, Cu/ZnSOD and actin as a loading control. **d**, MnSOD expression in wild-type or p27^{kip1}-/- MEFs infected with control retrovirus (puro) or FOXO3a-A3

(O3a) 24 h after infection. **e**, 3T3-L6 cells were grown asynchronously or serum starved (SF, serum free) for 48 h, and stimulated with insulin (Ins) for 24 h (top) or 10 min (bottom), with or without a 30-min pretreatment with LY294002 (LY). Lysates were analysed for expression of MnSOD and actin (top) or for phosphorylated PKB (PKB-pS473) and FOXO3a (FOXO3a-pT32; bottom). **f**, 293T cells were transfected with dominant-negative mutants of FOXO3a (DN1 or DN2). Cells were left untreated or treated with LY294002 for 16 h. Expression of MnSOD and actin (loading control), and phosphorylation of PKB on Ser 473 and FOXO3a on Thr 32 were analysed by immunoblotting.

almost basal amounts (Fig. 2b). Again, LY294002 fully reversed the inhibition induced by insulin.

We analysed the effect of specific Forkhead activation on a promoter fragment of 3,340 base pairs (bp) of the human *SOD2* gene for MnSOD (pSODLUC-3340)¹⁶. Luciferase expression controlled by this fragment was increased 3–4-fold on cotransfection with the various Forkhead transcription factors in several cell lines (data not shown), and also after 16 h of 4-OHT addition to DL23 but not DLD-1 cells (Fig. 2c). No effect was seen on a similar reporter construct lacking the sequences upstream of the transcription start site (pSODLUC-1; Fig. 2c). We could identify one inverse FOXO-binding element (DBE)¹⁷ at position –1,249 (GTAAACAA; DBE1) and one suboptimal DBE at position –997 (TTGTTTAA; DBE2) in the human *SOD2* promoter (Fig. 2d). A single G to C substitution in the TTGTTT sequence of DBE2 had no effect on

FOXO3a-induced luciferase activity of the *SOD2* reporter construct, but a similar substitution in the AAACAA sequence of DBE1 completely abolished FOXO3a-mediated expression of the *SOD2* gene (Fig. 2d). Chromatin immunoprecipitation assays also showed that FOXO3a could bind *in vivo* to the MnSOD promoter region encompassing DBE1 (Fig. 2e). Binding had already occurred as early as 4 h after 4-OHT treatment, which indicates that binding precedes the increase in mRNA MnSOD as determined by northern blotting (Supplementary Information Fig. 2). Together, these results show that expression of MnSOD is controlled transcriptionally by FOXO3a binding directly to DBE1.

The availability of glucose may be essential for growth-factor- and PKB-mediated protection from apoptosis². Because Forkheads become active in the absence of PKB activity, we reasoned that the regulation of antioxidant capacity by Forkhead-induced MnSOD regulation may protect cells from apoptosis, particularly under conditions of low glucose availability. We therefore analysed the effect of glucose deprivation on mitochondrial membrane integrity, an early marker for apoptosis, by rhodamine-1,2,3 staining¹⁸. Various cell types including DLD-1 and MEFs, which were cultured in the presence of 8% fetal calf serum (FCS) to activate PKB but were deprived of glucose for 16 or 72 h, respectively, showed a clear loss of mitochondrial membrane integrity (Fig. 3a). Pretreating the cells with 10 mM *N*-acetyl-L-cysteine (NAC), which enhances the scavenging of oxygen radicals, largely prevented apoptosis induced by glucose deprivation. This indicates that, in agreement with other reports^{19–21}, mitochondrial damage induced by glucose deprivation proceeds through increased production of ROS. Increased ROS production by glucose deprivation was also confirmed by means of an *in vitro* fluorescent assay similar to the *in vivo* assay described in Fig. 1 (Supplementary Information Fig. 3). We also showed that this effect of glucose deprivation was not due to depletion of cellular ATP resulting from diminished activity of the trichloroacetic acid cycle, as addition of pyruvate could not prevent the induction of mitochondrial instability (Fig. 3a). Mitochondrial leakage correlated with the induction of apoptosis in these cells, as determined by the fraction of cells containing a sub-G1 DNA content (Fig. 3a). Notably, whereas 4-OHT treatment was ineffective in DLD-1 cells, activation of FOXO3a in DL23 cells, either by the addition of 4-OHT or by LY294002 treatment, efficiently protected the cells from apoptosis induced by glucose deprivation (Fig. 3b).

Increased cellular ROS can result in the activation of stress-activated kinases, including Jun amino-terminal kinase (JNK). Indeed, glucose deprivation induced activation of JNK and this activation was inhibited in part by NAC pretreatment (Fig. 3c). More notably, the activation of JNK induced by glucose deprivation was inhibited after activation of FOXO3a, which indicates that the increased tolerance against oxidative stress induced by FOXO3a activation observed in the experiments using rhodamine-1,2,3 staining is reflected by a reduction in the induction of JNK (Fig. 3c).

To show that MnSOD is involved in this Forkhead-induced survival signal, we infected MnSOD-deficient MEFs²² with a retrovirus expressing FOXO3a–A3. Similar to DLD-1 cells, wild-type MEF cells deprived of glucose also lost mitochondrial membrane integrity unless infected with FOXO3a–A3 (Fig. 3d). In *Sod2*^{–/–} cells, however, mitochondrial damage occurred irrespective of the presence of FOXO3a–A3. Similarly, whereas LY294002 pretreatment efficiently protected wild-type cells, it had much less effect on *Sod2*^{–/–} cells. Ectopic expression of MnSOD in *Sod2*^{–/–} cells in similar amounts to those observed after inducing Forkhead activity made these cells refractory to glucose deprivation. We confirmed these data by analysing cell survival after glucose deprivation. Wild-type cells infected with FOXO3a–A3 could survive prolonged glucose deprivation (up to 5 d), whereas control-infected and *Sod2*^{–/–} cells expressing FOXO3a–A3 died in this period (Supplementary Information Fig. 3). FOXO3a-mediated survival is not a consequence of a reduced glucose requirement of arrested or

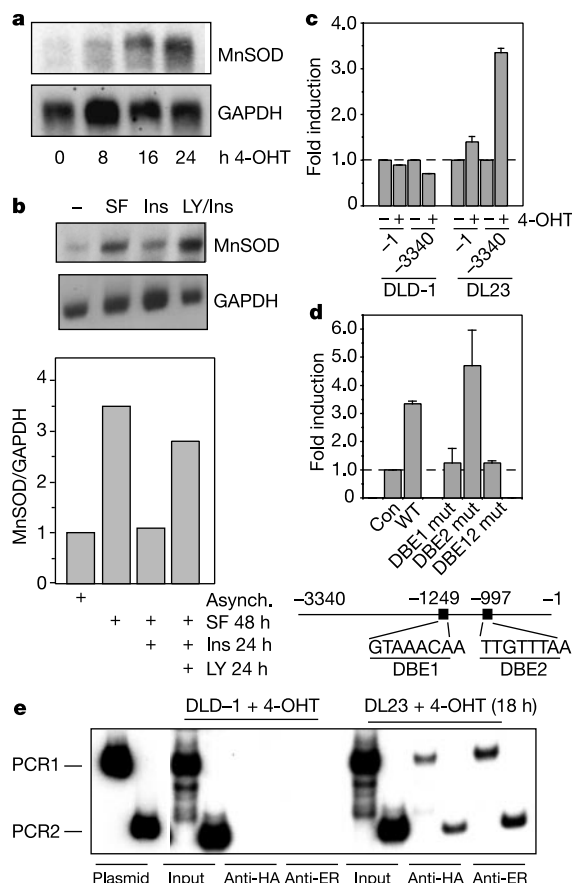


Figure 2 FOXO3a directly regulates the MnSOD promoter through an inverse DBE. **a**, mRNA (2 μ g) isolated from DL23 cells left untreated or treated with 500 nM 4-OHT was analysed by northern blotting for MnSOD and GAPDH (loading control). **b**, 3T3-L6 cells were treated as in Fig. 1e. Total RNA (30 μ g) was analysed for MnSOD and GAPDH mRNA by northern blotting. A representative experiment is shown; the histogram shows the mean of two independent experiments (SF, serum free; Ins, insulin; LY, LY294002). **c**, pSODLUC-3340 and pSODLUC-1 were transfected into DL23 or DLD-1 cells, which were left untreated or treated with 500 nM 4-OHT for 16 h before the luciferase activity was measured. **d**, pSODLUC-3340 carrying a point mutation at indicated positions was transfected into DL23 cells and luciferase activity was measured as in **c**. Data in **c** and **d** are the mean of three independent experiments. The linearized *SOD2* promoter fragment containing the two DBEs is also shown. DBE12mut, double mutant. **e**, DLD-1 and DL23 cells were treated with 4-OHT for 16 h, and then chromatin-bound DNA was isolated and immunoprecipitated with antibodies against the ER moiety or the HA epitope of HA-FOXO3a–ER. Immunoprecipitated DNA was analysed by PCR using two primer combinations, both encompassing DBE1. PCR was carried out on pSODLUC-3340 as a control for migration of the PCR fragments.

quiescent cells, because expression of the cell-cycle inhibitor p16 resulted in cell-cycle arrest but did not protect against glucose deprivation (Fig. 3d).

As reported previously^{10,23}, ectopic expression of FOXO3a reduces colony outgrowth of wild-type MEFs by inhibiting proliferation.

When we carried out the same experiment in *Sod2*^{-/-} cells, this resulted in the absence of any colony outgrowth (Fig. 3e), which indicates that cells entering quiescence through activation of Forkheads may require MnSOD for their survival. We therefore tested whether cells maintained in a quiescent state would show a similar

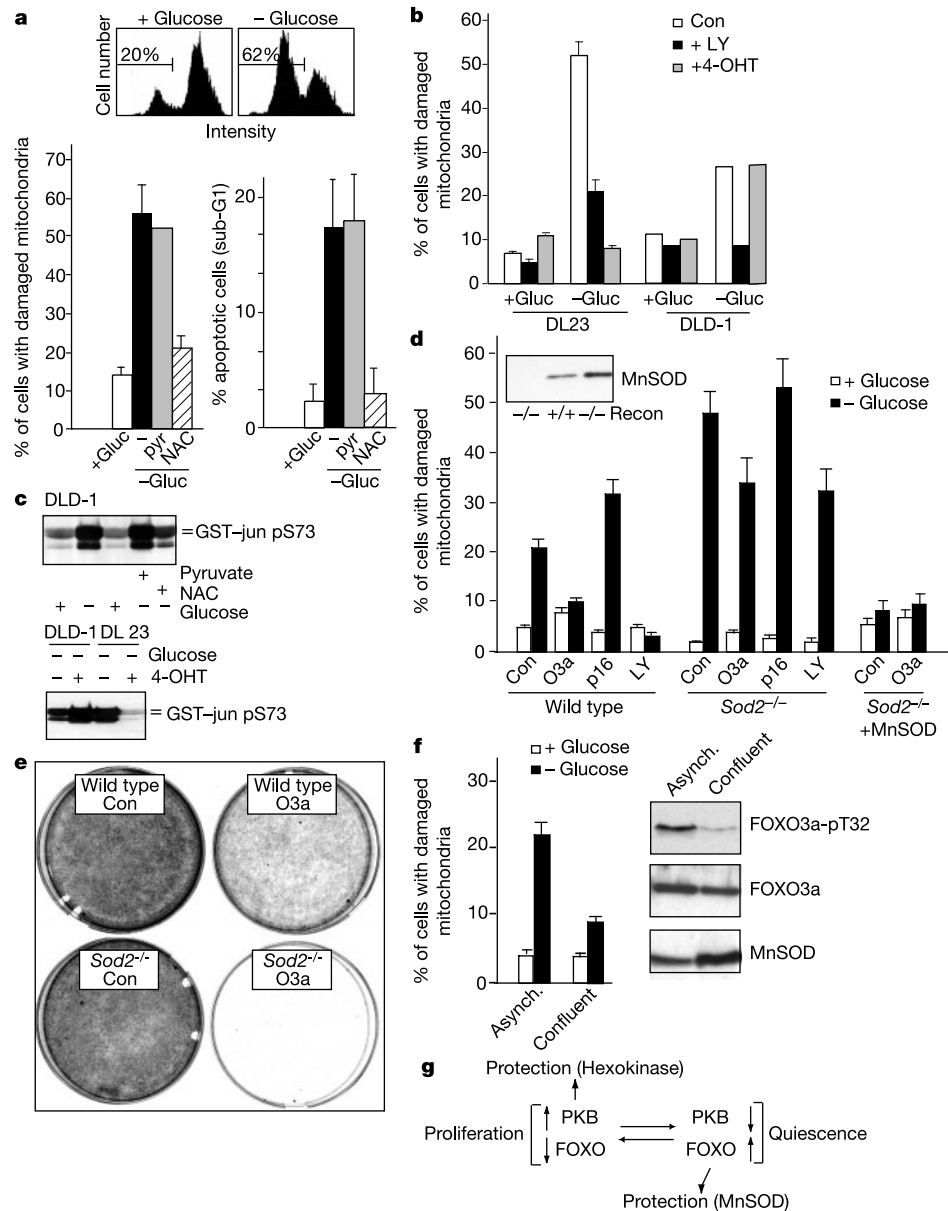


Figure 3 A Forkhead-mediated increase in MnSOD is required for the survival of arrested cells. **a**, DL23 cells were grown in DMEM medium supplemented with 8% FCS with or without glucose for 16 h, and mitochondrial membrane integrity was measured by rhodamine-1,2,3 staining. Typical cytometry profiles are shown on top. Left, DL23 cells were grown for 16 h with or without glucose, or without glucose and with pyruvate. Cells were left untreated or were treated with 10 mM NAC for 40 h (24 h before glucose withdrawal). Data are the means of at least three independent experiments. Right, cells were treated as above, but were fixed and stained with propidium iodide. The percentage of sub-G1 cells (representing apoptotic cells) was determined by flow cytometry. **b**, DLD-1 and DL23 cells were processed as in **a**. Cells were left untreated or were treated with LY294002 or 4-OHT for 40 h (24 h before glucose withdrawal). **c**, Cells were treated as indicated, and lysates were assayed for JNK activity. Treatment with 4-OHT, NAC and pyruvate was done as in **a** and **b**. **d**, Wild-type cells (+/+), *Sod2*^{-/-} cells (-/-) or *Sod2*^{-/-} cells with reconstituted MnSOD expression (-/- recon) were infected with a control virus (con) or with a virus expressing FOXO3a-A3 (O3a), and mitochondrial

membrane integrity was measured after 72 h of glucose withdrawal. Inset, quantity of MnSOD protein in the three cell lines. **e**, Wild-type or *Sod2*^{-/-} cells were infected as in **d**. After 2 weeks of selection, the cells were stained with crystal violet to visualize surviving colonies. **f**, Wild-type MEFs were grown to confluency, maintained for 4 d, and lysates were analysed for MnSOD and FOXO3a. Identical separate cultures were analysed by rhodamine staining for mitochondrial damage. **g**, Representation of the PKB-FOXO module, which functions as a molecular switch controlling cellular proliferation and protection from oxidative stress. Activation of PKB leads to inactivation of FOXO factors and cell-cycle entry. Simultaneously, cells become reliant on PKB-dependent survival signals that require the availability of glucose. Switching from active PKB to active FOXO factors will not only drive cells into a state of quiescence, but will replace simultaneously the PKB-dependent protection from oxidative insult by an alternative mechanism that no longer depends on glucose. This mechanism involves transcriptional upregulation of the MnSOD antioxidant enzyme.

resistance to glucose starvation by comparing contact-inhibited cultures with asynchronously growing MEFs. We found that contact-inhibited cells show increased protection against glucose deprivation, with a concomitant increase in their expression of MnSOD (Fig. 3f).

We have provided evidence that in the absence of PKB-mediated survival cells can be protected against oxidative damage by activating the Forkhead transcription factor FOXO3a, which directly increases the expression of MnSOD. These results show that the direct inverse coupling of Forkhead and PKB activity functions as a key switch in controlling the effects of cellular ROS. In a similar inverse manner, cell-cycle progression is controlled by PKB–Forkhead signalling. Activation of Forkheads results in entrance into a quiescent (G0) state¹¹, whereas inactivation of Forkheads by activation of PKB results in re-entry into the cell cycle and proliferation. Thus, regulation of proliferation and the appropriate protective mechanisms are coupled by the same PKB–Forkhead switch.

The dynamics of quiescence and proliferation set different requirements for the mechanism of ROS control, but also indicate why proliferation is coupled to glucose availability and why survival of quiescent cells must be relatively independent of their environment. This intricate link between proliferation or survival and glucose availability is also evident from results that we have obtained with tumour cells. In general, tumour cells are far more sensitive to glucose deprivation than are untransformed cells. In this respect, it is notable that tumour cells show high rates of aerobic glycolysis (the Warburg effect). In tumour cells, PKB is often constitutively activated either through deletions or mutations in PTEN (phosphatase and tensin homologue on chromosome 10) or through the activity of autocrine growth factors. Activation of PKB in tumour cells implies the inactivation of Forkhead and a loss of protection in the absence of glucose. Indeed, PTEN-negative U87MG glioblastoma cells show signs of substantial mitochondrial instability within a few hours of glucose deprivation (R.H.M., unpublished data), whereas MEFs show the same amount of mitochondrial damage only after prolonged deprivation (48–72 h). Others have reported an involvement of FOXO factors in oxidative stress responses¹⁴: expression of FOXO has been shown to increase cellular antioxidant activity, in keeping with the mechanism that we describe here. In contrast to that study, however, we did not observe any regulation of PKB activity (Supplementary Information Fig. 1) or FOXO transcriptional activity (data not shown) under conditions of low oxidative stress, such as glucose deprivation. In addition, whereas we have described a primary function of Forkheads in the defence against ROS (detoxification), others have shown that when ROS-induced DNA damage may occur Forkheads can contribute secondarily to defence by regulating the expression of proteins involved in the repair process²⁴.

In summary, we propose that the opposing action of PKB and the Forkhead transcription factors FOXO1, FOXO4 and FOXO3a on cell proliferation and their ability to provide protection against ROS-induced damage through differing mechanisms outlines a system that provides checks and balances for coupling cell-cycle regulation, glucose metabolism and protection from oxidative damage (Fig. 3g). In addition to the evolutionary conservation of the structure of this regulatory system in *Drosophila*, *C. elegans* and yeast (see refs 25, 26), we now show functional conservation as well. This indicates that the PKB–Forkhead pathway may be fundamental in coupling external cues to cellular response. □

Methods

General

Retroviral infections and maintenance of immortalized wild-type and p27^{kip1}−/− MEFs and primary wild-type and *Sod2*−/− MEFs have been described^{10,22}. The DL23 cell line has been described¹¹. We created the plasmid pBabe–FOXO3a–A3 by ligating a Klenow-blunted *HindIII*/*Bam*HI fragment of pcDNA3–HA–FOXO3a–A3 into Klenow-blunted pBabe–puro plasmid cut by *Bam*HI.

Chromatin immunoprecipitation assay

We carried out CHIP as described²⁷. About two 9-cm diameter dishes at 80% confluency were used per immunoprecipitation. Two sets of primers were designed to amplify by polymerase chain reaction (PCR) the region encompassing the FOXO3a-binding site in the *SOD2* promoter with the primers: PCR1, 5′-cagacgagctgtgtctc-3′ and 5′-gtccacgctgaattcc-3′; PCR2, 5′-ctagcttcggaagt-3′ and 5′-gttgccttattgaaag-3′.

Microscopy and fluorescence ratio imaging

We carried out microscopy and fluorescence ratio imaging essentially as described¹³. On oxidation, the fluorescence excitation/emission maxima of the C11-BODIPY^{581/591} probe shift from 581/591 to 490/510 nm, which facilitates measurement of the fraction of oxidized probe versus time. After adding 200 μM H₂O₂ to induce radical formation, we followed cells in time with dual excitation and emission confocal laser scanning microscopy. It should be noted that the C11-BODIPY^{581/591} probe is not oxidized by H₂O₂ itself, but by hydroxyl radicals formed in the Fenton reaction, H₂O₂ + O₂^{•−} → HO• + OH[−] + O₂, which is catalysed by divalent metal cations.

Measurement of mitochondrial membrane integrity

DL23 cells or MEFs were glucose deprived for 16 h or 72 h, respectively, by adding DMEM medium lacking glucose and pyruvate, but supplemented with 8% FCS. Note that owing to the small amount of glucose in FCS, glucose amounts are estimated to be about 10–20-fold less in this medium than in normal culture conditions. We added NAC (10 mM), LY294002 (20 μM) or 4-OHT (500 nM) 24 h before withdrawing glucose. The cells were then digested with trypsin and incubated with 10 μg ml^{−1} rhodamine-1,2,3 (ref. 18) at 37 °C for 30 min. The cells were washed twice with PBS and mitochondrial membrane staining was measured by standard flow cytometry.

JNK activity

After being washed in PBS, the cells were lysed in RIPA buffer and lysates were incubated with 5 μg of glutathione S-transferase (GST)-labelled Jun (residues 1–135) coupled to glutathione beads for 1 h at 4 °C. Beads were washed three times and incubated in kinase buffer (100 μM ATP, 20 mM MgCl₂; 50 mM Tris, pH 7.5; 1 mM DTT) for 30 min at 30 °C. We detected phosphorylation of GST–Jun by immunoblotting and probing with an antibody specific for Jun phosphorylated on Ser 73 (anti-pS73–Jun).

Western blotting and antibodies

Western blotting of total lysates was done as described⁸. Antibodies to MnSOD and Cu/ZnSOD were from StressGen. The antibodies to the ER (C-20) and actin (I-19) were from SantaCruz. The antibody against p27^{kip1} was from Transduction Labs. Antibody specific for PKB phosphorylated on Ser 473 (anti-pS473–PKB) and the anti-pS73–Jun antibody were from Cell Signaling. Antibody against FOXO3A phosphorylated on Thr 32 (anti-pT32–FOXO3A) was from Upstate Biotech. Before the immunoblotting shown in Fig. 1f, transfected cells were isolated by magnetic activated cell sorting procedure¹⁰.

Northern blotting

We ran 2 μg of mRNA (polyA-Tract, Promega) purified from 1 mg of total RNA (RNAzol, TEL-TEST; Fig. 2a) or 30 μg of total RNA (Fig. 2b) on a formaldehyde denaturing gel and transferred it to a GeneScreen-Plus nylon membrane (NEN). The blot was hybridized using radiolabelled probes for MnSOD (*Eco*RI fragment of pGEM3Z–MnSOD) and GAPDH (*Not*I-linearized pUC19–GAPDH). Plasmid pGEMZ–MnSOD was a gift from J. Wispe.

Luciferase assays

We carried out luciferase assays as described⁸. Plasmids pSODLUC-3340 and pSODLUC-1 (ref. 16) were a gift from M. Yim. We created pSODLUC-3340–DBE1mut and pSODLUC-3340–DBE2mut by site-directed mutagenesis using the primers 5′-ctgacgtctgaagaagccagcccttc-3′ and 5′-cattcaggattgtcttaactgttgag-3′, respectively.

Received 27 March; accepted 4 July 2002; doi:10.1038/nature01036.

1. Beckman, K. B. & Ames, B. N. The free radical theory of aging matures. *Physiol. Rev.* **78**, 547–581 (1998).
2. Gottlob, K. *et al.* Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes Dev.* **15**, 1406–1418 (2001).
3. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
4. Paradis, S. & Ruvkun, G. *Caenorhabditis elegans* Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. *Genes Dev.* **12**, 2488–2498 (1998).
5. Honda, Y. & Honda, S. The *daf-2* gene network for longevity regulates oxidative stress resistance and Mn-superoxide dismutase gene expression in *Caenorhabditis elegans*. *FASEB J.* **13**, 1385–1393 (1999).
6. Taub, J. *et al.* A cytosolic catalase is needed to extend adult lifespan in *C. elegans daf-C* and *clk-1* mutants. *Nature* **399**, 162–166 (1999).
7. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
8. Kops, G. J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
9. Rena, G., Guo, S., Cichy, S. C., Unterman, T. G. & Cohen, P. Phosphorylation of the transcription factor Forkhead family member FKHR by protein kinase B. *J. Biol. Chem.* **274**, 17179–17183 (1999).
10. Medema, R. H., Kops, G. J., Bos, J. L. & Burgering, B. M. AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27^{kip1}. *Nature* **404**, 782–787 (2000).
11. Kops, G. J. *et al.* Control of cell cycle exit and entry by protein kinase B-regulated forkhead transcription factors. *Mol. Cell. Biol.* **22**, 2025–2036 (2002).
12. Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. A modified oestrogen

- receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
13. Pap, E. H. *et al.* Ratio-fluorescence microscopy of lipid oxidation in living cells using C11-BODIPY^{581/591}. *FEBS Lett.* **453**, 278–282 (1999).
 14. Nemoto, S. & Finkel, T. Redox regulation of forkhead proteins through a p66^{shc}-dependent signaling pathway. *Science* **295**, 2450–2452 (2002).
 15. Dijkers, P. F. *et al.* FKHR-L1 can act as a critical effector of cell death induced by cytokine withdrawal: protein kinase B-enhanced cell survival through maintenance of mitochondrial integrity. *J. Cell Biol.* **156**, 531–542 (2002).
 16. Kim, H. P., Roe, J. H., Chock, P. B. & Yim, M. B. Transcriptional activation of the human manganese superoxide dismutase gene mediated by tetradecanoylphorbol acetate. *J. Biol. Chem.* **274**, 37455–37460 (1999).
 17. Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).
 18. Scaduto, R. C. Jr. & Grotyohann, L. W. Measurement of mitochondrial membrane potential using fluorescent rhodamine derivatives. *Biophys. J.* **76**, 469–477 (1999).
 19. Aulwurm, U. R. & Brand, K. A. Increased formation of reactive oxygen species due to glucose depletion in primary cultures of rat thymocytes inhibits proliferation. *Eur. J. Biochem.* **267**, 5693–5698 (2000).
 20. Lee, Y. J. *et al.* Glucose deprivation-induced cytotoxicity and alterations in mitogen-activated protein kinase activation are mediated by oxidative stress in multidrug-resistant human breast carcinoma cells. *J. Biol. Chem.* **273**, 5294–5299 (1998).
 21. Blackburn, R. V. *et al.* Metabolic oxidative stress activates signal transduction and gene expression during glucose deprivation in human tumour cells. *Free. Radic. Biol. Med.* **26**, 419–430 (1999).
 22. Huang, T. T. *et al.* Superoxide-mediated cytotoxicity in superoxide dismutase-deficient fetal fibroblasts. *Arch. Biochem. Biophys.* **344**, 424–432 (1997).
 23. Nakamura, N. *et al.* Forkhead transcription factors are critical effectors of cell death and cell cycle arrest downstream of PTEN. *Mol. Cell. Biol.* **20**, 8969–8982 (2000).
 24. Tran, H. *et al.* DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein. *Science* **296**, 530–534 (2002).
 25. Kenyon, C. A conserved regulatory system for aging. *Cell* **105**, 165–168 (2001).
 26. Guarente, L. & Kenyon, C. Genetic pathways that regulate ageing in model organisms. *Nature* **408**, 255–262 (2000).
 27. Boyd, K. E., Wells, J., Gutman, J., Bartley, S. M. & Farnham, P. J. c-Myc target gene specificity is determined by a post-DNA binding mechanism. *Proc. Natl Acad. Sci. USA* **95**, 13887–13892 (1998).
- Supplementary Information** accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).
- Acknowledgements**
We thank P. Dijkers, M. Yim, D. Powell and J. Wispe for reagents; W. R. Sellers for discussions; M. Daniels for technical assistance; and C. Marshall for critically reading the manuscript. G.J.P.L.K. is supported by Chemical Sciences, T.T.H. is supported by the NIH.
- Competing interests statement**
The authors declare that they have no competing financial interests.
Correspondence and requests for materials should be addressed to B.M.T.B. (e-mail: b.m.t.burgering@med.uu.nl).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Taking the initiative

Biomedical research in London sprawls throughout the city, spread between various university colleges, hospitals and numerous biotechnology companies (see pages 4–5). A prime example of the sort of components that make up this disparate network is the National Institute for Medical Research in Mill Hill, north London, which employs about 600 people.

The institute is run by Britain's Medical Research Council (MRC), and is seen by many as a desirable place to work because it has steady funding, good facilities and is reasonably close to potential collaborators with expertise in other disciplines. In addition, because the MRC is so closely involved, research at the institute is somewhat less susceptible to the political vagaries that make funding of the council's off-site projects a little less stable (see *Nature* 418, 714; 2002).

But these attributes do have a downside — they make it a lot harder to secure a post there. So how would one go about landing a postdoc position in a lab such as this? The usual way is simply by responding to an advertisement, says Steve Ley, an immunologist at the institute. But a more effective approach might be to find out who at a lab is doing the kind of work you want to pursue, reading up on their work and then writing them a letter regardless of whether a position has actually been advertised.

A French graduate student, motivated by the lack of academic postdoc positions in her own country, landed a slot in Ley's lab using just this method. "It shows a level of initiative that you might not get from adverts," Ley says. And with increasing international competition — 75% of the Mill Hill facility's postdocs are from abroad — that level of initiative may well prove to be necessary.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

London: a question
of space

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS





Calling for entrepreneurs London

One spin-off a month, one patent a week: that is the technology-transfer goal trumpeted by the University of London's Imperial College of Science, Technology and Medicine. And the science departments of two other major London colleges — King's and University College London (UCL) — are taking up a similar battle cry.

The three colleges' increasingly active technology-transfer offices (see 'Transferring ideas', opposite) are helping to provide the impetus for the creation of new companies and job opportunities within the capital. "There's a huge engine for innovation here," says Leszek Borsiewicz, principal of Imperial's faculty of medicine.

But the question remains whether that engine can power London to overtake rivals Oxford and Cambridge — especially as these two cities have the one resource London lacks: space. Borsiewicz notes that the University of London receives more investment from government and industry than either of the rival regions. But Oxford and Cambridge each boast a more developed biotech climate and fewer infrastructure problems.

Access to London's financial resources, business-development community and legal expertise is aiding the colleges in their new role as business incubators. "There's masses of financial resources here. There's private equity here. Most of the venture capitalists have an office here," says Reading-based William Powlett Smith, a partner with financial analysts Ernst & Young.

But problems such as the high cost of living and difficulties involved in getting into and around the city threaten to choke business development. "You can get around easily — if you travel at 4 a.m.," says Smith. And without space for companies to grow into, the engine could be all revved up with nowhere to go. But the colleges and companies emerging from them are taking creative approaches to confront those problems.

ACADEMIC SPARKPLUGS

Richard Sykes, former non-executive chairman of drug firm GlaxoSmithKline, took over as rector of Imperial in 2001. He is keen to promote an atmosphere of entrepreneurship at his new home. "There is a pervasive culture at Imperial where technology transfer is seen as an important part of your mission," Borsiewicz says. Sykes has also made starting salaries more attractive to new recruits, and has publicly called for more funding to allow for higher salaries.

Higher salaries would certainly help attract people to what Arthur Lucas, professor of science curriculum studies at King's College, calls an "odd" labour market. "London is one of the most expensive cities in the world," he says. "On the other hand we've been very successful in attracting people from other parts of the world, including the east and west coasts of the United States." For Lucas, London's attractions include access to museums, libraries and organizations such as the Royal Society. It also has strong transport links both by air and rail.

UCL has also found ways to attract new recruits. For example, it has expanded its Institute of Ophthalmology thanks to £8.8 million (US\$13.7 million) from the Wellcome Trust, Britain's largest biomedical

Making capital: Imperial College is one of three London colleges seeking to spin off companies from its successful research.



research foundation, and eye-research charity Fight for Sight, together with £6.5 million from drug firm GlaxoSmithKline (GSK). The immunology chair set up by GSK lured Santa Ono to the institute from Harvard University. Ono says the expanded institute represents “a pretty remarkable package”, noting that the funding and infrastructure have already enticed a dozen other non-UK scientists to the institute.

Many scientists who already live and work in the London don't want to leave, despite the city's urban hassles. As a result, companies have looked for ways to stay in the capital rather than relocate to other cities' science parks. Ken Powell, chief executive of biotech company Arrow Therapeutics, says that many employees bristled when some venture capitalists who invested in his company suggested that it should move to Cambridge or Oxford. “We would have lost half of our key staff,” Powell says. “They wanted to be in London.”

So Powell and colleagues looked for a building in the city. They found a dishevelled office building south of the River Thames and persuaded the landlord to cover half of the costs involved in converting it into a lab. They now pay a lower rent than they would in the science parks of Cambridge or Oxford.

STAYING PUT

But as about 70% of London's 80 or so biotech companies have no more than 10 employees, and less money behind them than Arrow has, they are not in the market for converting office buildings. Peter Shepherd, chief executive of Xcellsys, a UCL spin-off, has solved this problem by establishing a ‘virtual’ corporate presence in a building that rents offices and meeting rooms out by the day, week or hour. The arrangement allows him to separate his company business from his academic research, and helps him to capitalize on the easy access to venture capitalists, lawyers and accountants that his location affords.

A more common arrangement in London is to rent space back from the institution that created the company. Aeres Biomedical, which emerged from the antibody-engineering programme run by the Medical Research Council (MRC), occupies a building owned by the MRC near its Mill Hill campus in north London. Tarran Jones, Aeres's chief executive, says that he would like the company to stay in London but space at the MRC institute is getting tight. But moving too far away would put the company at risk of losing the expertise that has allowed it to advance. Even moving across the Thames, like Arrow, could be problematic. North and south London are “worlds away”, Jones says.

A FINE SPREAD

The large distances between companies are characteristic of the London biotech scene and set it apart from other regions where activity tends to be more concentrated. “There's no such thing in my mind as a London biotechnology cluster,” says Anthony Walker, chief executive of Onyvax, which works on cancer vaccines. This company rents space from St George's Hospital Medical School in south London and is opening a satellite office at the Royal Veterinary College in central London, which funds the London BioScience Innovation Centre, the capital's only biotech incubator. But Onyvax

Transferring ideas

Perhaps the best sign of how London's research climate is turning entrepreneurial is in the growth of the University of London's technology-transfer offices.

The biggest of the three, Imperial College Innovations, has four divisions and 23 staff dedicated — up from eight five years ago. Innovations helps fledgling companies to draw up business plans, funds patent applications, puts up as much as £250,000 (US\$388,000) in seed funds and locates other investors.

UCL Ventures, University College London's tech-transfer arm, has added four specialists to its office

within the past six months, bringing the total to 10. UCL Ventures differs from Imperial's version in that it only incorporates companies that have received investment. But it has access to one seed fund, as well as a network of venture capitalists.

KCL Enterprises, which serves King's College, has seen perhaps the most dramatic growth, from a one-person office to a staff of six over five years. It helps young companies to grow by deploying two seed funds, which are supplied primarily by the Wellcome Trust and the UK government.

P.S.

is seeking still more space, because it has a therapy for prostate cancer in phase II trials.

Walker jokes with Glyn Edwards, chief executive of biotech firm Antisoma, about taking over his extra laboratories. “We've got spare space,” Edwards smiles. “But if we let them in, we won't have anywhere to expand into.”

Edwards says that jostling for space is the price a company pays for staying in London. For Antisoma, which seeks to develop cancer therapies discovered in academic labs, that price is worthwhile. Although the company looks around the world for proteins and small molecules, London connections often lead the way. Networking with London-based Cancer Research UK helped Antisoma to secure development rights to a platinum-based chemotherapy molecule developed by scientists at the German Cancer Research Center in Heidelberg.

Not all companies have problems claiming space for themselves. Businesses based on bioinformatics software, databases and data mining have managed to stay central. Inpharmatica is one such example. “We're almost at the centre of the Underground map,” says John Overington, vice-president of drug discovery. But as the company moves further into drug discovery it will need to develop lab space in another location.

Identifying and developing such locations may be key to London's ability to attract academics who also want to be entrepreneurs — and to expand jobs for scientists hoping to move into industry. Perhaps a strong network of colleges, hospitals and companies may be able to attract the government support necessary to solve the space issue. Because without room to grow, companies will be driven away to other regions.

Paul Smaglik is editor of *Naturejobs*.